



HAL
open science

L'effet des inégalités d'éducation sur le développement économique : un essai d'évaluation

Wail Benaabdelaali

► To cite this version:

Wail Benaabdelaali. L'effet des inégalités d'éducation sur le développement économique : un essai d'évaluation. Economies et finances. Université de Toulon, 2017. Français. NNT : 2017TOUL2007 . tel-01817640

HAL Id: tel-01817640

<https://theses.hal.science/tel-01817640>

Submitted on 18 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'effet des inégalités d'éducation sur le développement économique : un essai d'évaluation

Thèse de doctorat en Sciences économiques présentée par

Wail BENAABDELAALI

soutenue le 20/10/2017 à l'Université de Toulon, Campus Porte d'Italie

Directeur de thèse : **Maurice CATIN**
Co-encadrant de thèse : **Saïd HANCHANE**

Jury :	Jean BOURDON <i>Rapporteur</i>	Directeur de recherche émérite au CNRS, IREDU, Université de Bourgogne
	Maurice CATIN <i>Directeur de thèse</i>	Professeur émérite, LÉAD, Université de Toulon
	Michaël GOUJON <i>Rapporteur</i>	Maître de conférences HDR, CERDI, Université d'Auvergne
	Saïd HANCHANE <i>Co-directeur de thèse</i>	Professeur, Université Polytechnique Mohamed VI, EGE de Rabat ; HDR, Université de Toulon
	Christophe MULLER <i>Suffragant</i>	Professeur, GREQAM, Aix-Marseille Université

L'université de Toulon n'entend donner aucune approbation ou improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

Remerciements

Cette thèse de doctorat n'a pu devenir réalité sans le soutien et la forte implication de nombreuses personnes qui m'ont épaulé tout au long de ces années de recherche. A ce titre, je souhaite aujourd'hui leur adresser mes plus vifs remerciements.

En premier lieu, je souhaite exprimer ma plus profonde reconnaissance et toute ma sympathie à Monsieur le Professeur Maurice Catin qui a accepté de diriger cette thèse. Son encadrement est avant tout un honneur. Ses remarques et orientations tant sur le plan analytique que méthodologique ont été indispensables et déterminantes pour améliorer la qualité du travail accompli.

Ma sincère gratitude va également à mon codirecteur de thèse, Monsieur le Professeur Said Hanchane qui a été aussi mon ancien directeur à l'Instance Nationale d'Evaluation (INE) entre 2009 et 2014. Son enthousiasme scientifique a suscité mon intérêt pour la voie doctorale après le diplôme d'ingénieur. Son soutien scientifique, méthodologique et parfois psychologique m'ont été très précieux.

Je suis très sensible à l'honneur que m'ont fait Messieurs Jean Bourdon et Michaël Goujon d'avoir accepté d'évaluer mon travail en tant que rapporteurs. Je voudrais également exprimer ma reconnaissance au Professeur Christophe Muller pour sa présence au sein du jury.

Que soient aussi chaleureusement remercié tout le personnel de l'école doctorale, les membres du LÉAD, mes amis et collègues de l'Université de Toulon. Mes remerciements vont également à Abdelhak Kamal pour m'avoir accompagné et encouragé pendant toute la durée de ce travail.

Je voudrais également remercier toutes les personnes attentives à l'aboutissement de cette thèse. Ils sont très nombreux pour que je puisse les nommer mais qu'ils soient tous assurés de ma sincère gratitude.

Ces remerciements seraient incomplets sans mentionner mes collègues de tous les jours de l'Instance National d'Evaluation en particulier et du Conseil supérieur de l'éducation en général, avec qui j'ai partagé plus de neuf années de ma vie.

Et évidemment... mes chers parents, ma femme, mon frère, mes enfants, ma famille, vous qui m'avez accordé votre affectueuse bienveillance, précieux soutien et à qui j'ai dû manquer, cette thèse est un peu la vôtre.

A ma famille

Sommaire

Introduction générale	11
Chapitre I. La mesure des inégalités éducatives	23
1.1. Introduction	23
1.2. Comment mesurer les inégalités dans l'éducation ?.....	24
<i>1.2.1. La construction des indices d'inégalités éducatives : cadre méthodologique</i>	<i>24</i>
<i>1.2.2. Les indices d'inégalités éducatives : revue de littérature</i>	<i>28</i>
<i>1.2.3. Décomposition des indices d'inégalités éducatives</i>	<i>32</i>
<i>1.2.4. Les bases mondiales des inégalités éducatives existantes</i>	<i>36</i>
1.3. Proposition d'une nouvelle base mondiale sur les inégalités éducatives	39
<i>1.3.1. Méthodologie</i>	<i>40</i>
<i>1.3.2. Calcul des durées des cycles d'enseignement sur la période 1950- 2010</i>	<i>50</i>
1.4. Relations entre l'indice de Gini de l'éducation, le taux d'analphabétisme et la moyenne d'années de scolarisation.....	52
<i>1.4.1. L'indice de Gini de l'éducation et le taux d'analphabétisme.....</i>	<i>52</i>
<i>1.4.2. Le lien entre l'indice de Gini de l'éducation et le niveau moyen d'éducation: vers une formulation générale</i>	<i>54</i>
<i>1.4.3. Le domaine de variation de l'indice de Gini de l'éducation selon la moyenne d'années de scolarisation</i>	<i>59</i>
1.5. Tendances des inégalités éducatives selon les niveaux de développement des pays	62
1.6. Conclusion	70
Chapitre II. Inégalités éducatives et développement : l'existence d'une relation non linéaire.....	74
2.1. Introduction	74
2.2. Les enseignements de la revue de la littérature	75
2.3. La spécification non linéaire du modèle	83
<i>2.3.1. Les variables retenues</i>	<i>83</i>

2.3.2.	<i>La spécification du modèle estimé</i>	85
2.4.	L'impact des inégalités éducatives sur le développement économique : une vérification empirique de la non-linéarité	90
2.4.1.	<i>Modèle non linéaire polynomial</i>	91
2.4.2.	<i>Modèles non-paramétrique et semi-paramétrique</i>	95
2.4.3.	<i>Décomposition inter et intra-groupes des inégalités et contribution au développement économique</i>	103
2.5.	Conclusion	107
Chapitre III. Les effets des inégalités éducatives sur le développement territorial : le cas du Maroc		
3.1.	Introduction	111
3.2.	Inégalités éducatives et développement humain dans les régions marocaines	111
3.2.1.	<i>IDH et indice de Gini de l'éducation : revue de littérature et méthodologie</i>	111
3.2.2.	<i>Les inégalités régionales d'éducation au Maroc 1982-2010</i>	117
3.2.2.1.	<i>Profil régional des inégalités d'éducation : un lent processus de convergence</i>	117
3.2.2.2.	<i>Des inégalités de genre et de milieu plus prononcées dans les régions les moins développées</i>	121
3.2.3.	<i>Les disparités régionales de développement au Maroc</i>	125
3.2.3.1.	<i>Le niveau de développement humain des régions marocaines</i>	125
3.2.3.2.	<i>Les écarts de développement entre régions</i>	129
3.3.	Inégalités éducatives et développement humain au niveau communal : une approche par l'économétrie spatiale	132
3.3.1.	<i>Rappel méthodologique</i>	132
3.3.2.	<i>Inégalités spatiales d'éducation et de développement humain au Maroc : approche par l'analyse exploratoire des données spatiales</i>	135
3.3.3.	<i>Les effets des inégalités d'éducation sur le développement humain</i>	141
3.3.3.1.	<i>Définition des variables</i>	141
3.3.3.2.	<i>Un modèle économétrique spatial</i>	143
3.3.3.3.	<i>Un modèle polynomial spatial</i>	144
3.4.	Inégalités éducatives et développement économique des provinces marocaines	147

<i>3.4.1. Les enseignements de la littérature au niveau territorial</i>	147
<i>3.4.2. Définition des variables et spécification du modèle</i>	148
<i>3.4.3. Résultats du modèle</i>	149
<i>3.4.4. Décomposition inter et intra-groupes des inégalités et contribution au développement économique des provinces marocaines</i>	153
3.5. Conclusion	156
Conclusion générale	161
Annexes	169
Bibliographie	179
Index des tableaux	195
Index des graphiques	197
Index des figures	199

Introduction générale

1. **Le rôle du capital humain dans la croissance économique**
2. **De l'inégalité de revenu aux inégalités du capital humain dans le processus de croissance économique**
3. **Apports et problématique de la thèse**
4. **Plan de la thèse**

Introduction générale

Les liens entre capital humain, inégalités et développement sont complexes et ont fait l'objet de nombreux travaux aussi bien théoriques qu'empiriques. Les prix Nobel¹ décernés en faveur de ces champs d'analyse ont montré l'importance qu'on leur confère. Le consensus aujourd'hui en ce qui concerne les bienfaits de l'éducation a amené l'adoption par la communauté internationale de plusieurs déclarations et conventions sur le « droit à l'éducation » et « le droit à l'égalité »². Cette préoccupation mondiale face à ces droits fondamentaux pour chaque individu, qui figurait parmi les objectifs du millénaire du développement 2000-2015, revient de nouveau à l'ordre du jour dans le nouvel agenda international. Le programme de développement durable³ 2015-2030 réitère, dans les quatrième et dixième objectifs, l'importance de réduire les inégalités au sein et entre les pays et d'assurer, à tous les individus, un accès équitable à une éducation de qualité.

Les relations entre éducation, inégalités et développement ont suscité l'intérêt de plusieurs disciplines autres que l'économie. Mais, c'est particulièrement dans le champ de l'économie du développement que ces relations n'ont cessé de connaître des regains d'intérêt sur le plan analytique, que ce soit après la relance du débat sur la convergence des économies ou suite à l'émergence de la théorie de la croissance endogène.

1. Le rôle du capital humain dans la croissance économique

- *Du modèle néoclassique à la théorie de la croissance endogène*

Plusieurs travaux ont tenté d'évaluer le rôle du capital humain dans la croissance économique. Les premiers travaux ont pour origine la version étendue du modèle de croissance néoclassique de Solow (modèle de Solow augmenté, Mankiw et al. 1992). Les estimations issues de ce modèle affirment l'effet positif de l'éducation sur la croissance. Néanmoins, on s'est rapidement aperçu que ces estimations reposent sur un grand nombre d'hypothèses réductrices relatives au capital humain. En effet, les modèles néoclassiques considèrent le capital humain comme toute autre forme de capital « classique » qui permet de procurer un revenu à travers l'investissement et qui peut aussi s'accumuler ou s'user. Néanmoins, à la différence du capital physique, le capital humain ne peut être parfaitement échangé, subdivisé ou transféré entre les individus. Il est indissociable et totalement incorporé à son détenteur.

Une deuxième série de travaux issus de la théorie de la croissance endogène distingue l'accumulation du capital humain de l'accumulation du capital physique. De tels travaux soutiennent plutôt le caractère « auto-entretenu » de la croissance par le biais de l'éducation et prennent en considération ses éventuelles externalités.

¹ Simon Kuznets (Prix Nobel d'économie 1971), Theodore Schultz (Prix Nobel d'économie 1979), Gary Becker (Prix Nobel d'économie 1992) et Joseph Eugene Stiglitz (Prix Nobel d'économie 2001).

² On peut citer notamment la Déclaration Universelle des Droits de l'Homme de 1948 (article 26) et la Convention Internationale relative aux Droits de l'Enfant de 1989 (article 28).

³ Le nouveau programme de développement durable a été adopté par l'assemblée générale des Nations Unies le 25 septembre 2015

Contrairement aux modèles de croissance néoclassiques, le stock de capital humain n'est plus considéré comme étant relié aux caractéristiques exogènes de l'économie et la croissance économique peut, dans ce cas, être soutenue sur une longue période. Pour les modèles de croissance néoclassiques, la convergence des pays est principalement due à la tendance de ces derniers à enregistrer des rendements décroissants au fur et à mesure qu'ils tendent vers des niveaux plus élevés de développement. Ainsi, à travers le capital humain, la théorie de la croissance endogène apporte une explication aux différents processus possibles de convergence ou de divergence des économies.

Les modèles de croissance endogène attachent une priorité particulière à l'accumulation du capital humain comme un préalable au développement technologique. Contrairement aux modèles néoclassiques dans lesquels le flux du capital humain est plus important, les variables de stock du capital humain dans les modèles de croissance endogène permettent de relâcher l'hypothèse de la proximité de l'équilibre stationnaire. L'estimation directe de la fonction de production agrégée conduit à des résultats plus ouverts puisqu'elle permet de contourner des hypothèses peu vraisemblables sur la nature de l'équilibre.

- ***Capital humain et croissance : le questionnement de la littérature empirique***

Sur le plan statistique, la construction de bases comparables au niveau international utilisant des données de flux⁴ est relativement moins complexe que celle procurant des données de stock. C'est pour cette raison que le premier panel de données de stock n'a été réalisé qu'en 1991 par Kyriacou (1991) et le premier panel cylindré (complet d'observations) n'a été disponible qu'en 1993 par Barro et Lee (1993). Cependant la première base de données concernant la distribution de ce stock de capital humain n'a pu être élaborée qu'en 2001 par Thomas et al. (2001) et mieux détaillée par Benaabdelaali et al. (2012).

Suite à la production de bases avec l'usage des données de stock, les résultats des estimations de la relation entre capital humain et croissance semblent différents. Avec la spécification basée sur la fonction de production, Benhabib et Spiegel (1994)⁵ et Pritchett (1996)⁶ trouvent des effets soit non significatifs soit négatifs. Avec une spécification issue du modèle de convergence, l'effet du stock du capital humain sur la croissance apparaît positif et significatif. En introduisant directement le stock de capital humain issu de la base de Barro et Lee (1993) dans l'équation de convergence conditionnelle, Islam (1995) rejoint les résultats trouvés par Mankiw et al. (1992) où le coefficient du capital humain est significativement positif.

A première vue, il semblerait donc que ces résultats divergents dépendent largement de la spécification retenue. Néanmoins, Gurgand (2000) souligne que même lorsque les méthodes économétriques les plus robustes sont employées, l'effet positif et significatif peut s'avérer inexistant, quelle que soit la spécification économique retenue. Plus surprenant encore, des analyses montrent que le stock de

⁴ Ces données de flux concernent principalement les taux de scolarisation.

⁵ Benhabib et Spiegel (1994) utilisent la base de Kyriacou (1991).

⁶ Pritchett (1996) utilise la base de Barro et Lee (1993) et de Nehru et al. (1995).

capital humain impacte négativement la croissance. Ces conclusions concernent aussi bien les anciens que les récents travaux sur cette relation. Benos et Zotou (2014), qui ont dressé une méta-régression⁷ combinant les résultats de 989 régressions sur le lien entre éducation et croissance, découlant de 57 études menées sur données transnationales, confirment que les conclusions demeurent le plus souvent ambiguës et non convergentes.

C'est ainsi que de nouveaux questionnements sur les hypothèses qui sous-tendent la relation capital humain/croissance ont émergé. Comment expliquer que cette relation soit vérifiée au niveau micro-économique alors qu'elle ne l'est pas toujours au niveau macro-économique ? Plusieurs économistes se sont posés la question de savoir pourquoi ce qui est bon pour les individus ne l'est pas pour la nation⁸. Certains se sont même posés la question "*where has all the education gone?*" (Pritchett, 2001; Ahsan et Haque, 2017).

2. De l'inégalité de revenu aux inégalités du capital humain dans le processus de croissance économique

Tout comme la relation entre capital humain et croissance, celle entre inégalités et croissance est aussi largement débattue dans les modèles de croissance. Le premier modèle théorique établi sur cette relation (Kaldor, 1956), s'inscrit dans la lignée de la théorie classique selon laquelle les inégalités sont une condition préalable à la croissance économique. Dans les années 1970 et 1980, malgré l'intérêt des économistes porté sur la relation de causalité inverse, l'idée conventionnelle répandue était que l'élargissement des inégalités n'était pas nuisible à la croissance. Au contraire, selon la théorie classique, une répartition inégalitaire des revenus favorise au départ le flux d'épargne qui peut stimuler la croissance économique.

Le manque de données comparables au niveau international sur les inégalités de revenu a constitué l'obstacle principal à la validation de la relation. Ce n'est qu'à partir des années 1990 que certains travaux ont tenté de vérifier empiriquement la relation moyennant des données transversales. Les résultats de ces travaux vont faire émerger une nouvelle vision qui révèle plutôt un lien négatif entre les inégalités « initiales » de revenu et la croissance de long terme (Alesina et Rodrik 1994; Person et Tabellini, 1994; Clarke, 1995).

Cependant, suite à la publication de la première base mondiale de large couverture sur les inégalités de revenu (Deininger et Squire, 1996), la nature de la relation va être de nouveau remise en question. Les résultats de la modélisation en données de panel permettant de contrôler les effets fixes « pays », suggèrent plutôt une relation positive à court et à moyen terme (Li et Zou, 1998 ; Forbes, 2000).

Mais durant ces dernières années, de nouveaux travaux, basés aussi sur des données de panel, aboutissent à des conclusions différentes. Ainsi, suite à la construction de nouvelles bases mondiales des inégalités de revenu (WIIDv1, WIIDv2c, SWIIDv1.0, SWIIDv2.0), Castelló (2010a), Cingano (2014), Halter et

⁷ La méta-régression est une approche de la méta-analyse qui consiste à utiliser les résultats d'une série de régressions issues d'études indépendantes et disparates afin de faire une synthèse objective.

⁸ Dans la mesure où les estimations issues de fonctions de gains de Mincer aboutissent généralement à l'identification de taux de rendement positifs.

al. (2014) et Ostry et al. (2014) montrent que les inégalités ont plutôt des effets néfastes sur la croissance économique dans les pays en développement.

De cette abondante littérature sur la relation inégalités/croissance économique, il faut donc souligner que la question de savoir si les inégalités sont nuisibles ou favorables au développement économique n'est pas tranchée. La polémique sur cette relation est toujours d'actualité et demeure sujette à un re-questionnement permanent principalement après la publication de nouvelles bases de données.

- ***La prise en compte des inégalités en matière d'éducation dans les liens éducation/croissance et inégalités/croissance***

Parmi les explications plausibles qui peuvent être avancées pour justifier l'absence de consensus sur les relations capital humain/croissance et inégalités/croissance, figure les problèmes de mesure. En effet, la majorité des travaux qui ont analysé la relation entre capital humain et développement économique ont principalement appréhendé le capital humain à travers des mesures quantitatives de l'éducation. "However, lack of data have led to serious limitations: educational variables such as schooling rates or the average number of school years are vague indicators of the measurement of human capital relating to education." (Altinok, 2007).

De même pour l'analyse du lien inégalités/croissance, la variable d'inégalité a été le plus souvent traitée d'un point de vue exclusivement monétaire en se basant sur « le revenu » et dans une moindre mesure sur « la propriété foncière ». "However, income and land inequality may be *insufficient measures of wealth inequality* since other variables such as human capital are also important determinants of wealth and growth" (Castelló et Doménech, 2002).

Ainsi, plusieurs travaux sur les deux relations précédemment citées ont soulevé le manque de données comparables au niveau mondial sur des indicateurs non quantitatifs du capital humain (Altinok, 2007) ou sur des mesures non économiques des inégalités. "...the measurement of non-economic inequality has been held back for a long time by lack of data" (Peterson, 2014). "Due to the relative lack of available data on educational inequality, the impact of inequality on growth remains underexplored from an empirical perspective" (Rodríguez-Pose et Tselios, 2010). Dessus (2000) signale que les données statistiques dont nous disposons jusqu'à présent ne permettent pas de construire une mesure satisfaisante des inégalités éducatives.

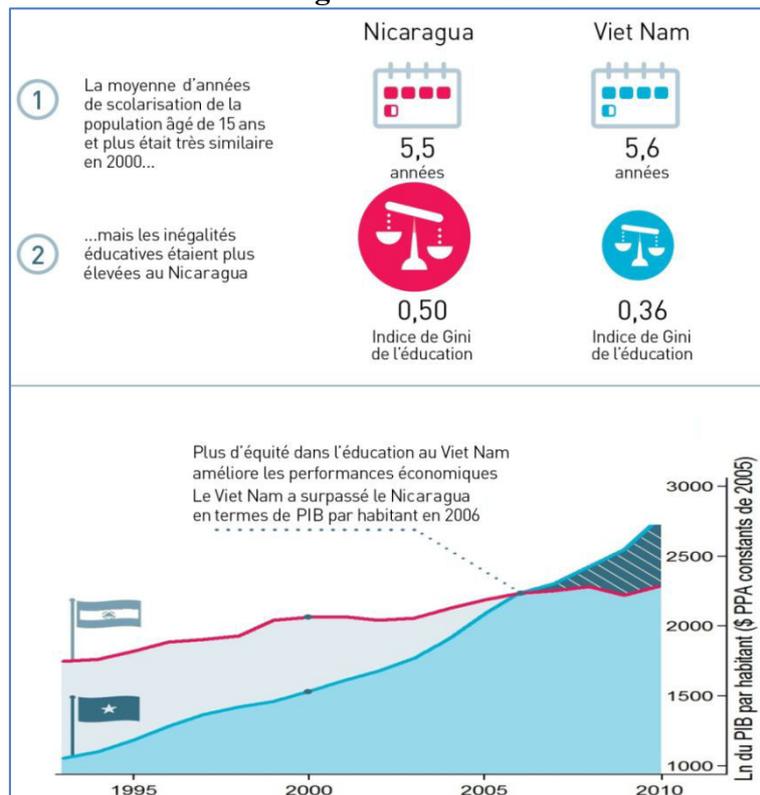
Par conséquent, la variable de « l'inégalité en matière d'éducation » est considérée comme une mesure alternative à la fois de l'inégalité de revenu dans le lien inégalités/croissance et du stock de capital humain dans la relation éducation/croissance. Or deux pays enregistrant le même niveau d'inégalité de revenu peuvent néanmoins connaître des taux de croissance économique différents si les niveaux de l'inégalité d'éducation de ces deux pays sont dissemblables.

De même, deux pays qui disposent de stocks de capital humain assez similaires peuvent toutefois présenter des tendances différentes de développement économique si l'un des deux présente une distribution de l'éducation relativement

plus asymétrique que celle de l'autre. Le graphique n°1 illustre ce cas à travers une comparaison entre deux pays : le Nicaragua et le Vietnam.

Le stock de capital humain mesuré par la moyenne d'années de scolarisation de la population âgée de 15 ans et plus était, en 2000, très similaire pour ces deux pays : 5,5 années au Nicaragua et 5,6 années au Vietnam. Néanmoins, le niveau d'inégalité d'éducation mesuré par l'indice de Gini de l'éducation au Nicaragua était supérieur de 40% à celui du Vietnam pour cette même année. Et le niveau plus élevé d'inégalité d'éducation au Nicaragua semble ralentir la croissance dans ce pays (graphique n°1).

Graphique 1 : Inégalité d'éducation et développement économique, cas du Nicaragua et du Vietnam



Source : Auteur.

3. Apports et problématique de la thèse

Les analyses empiriques, menées sur plus d'un demi-siècle, aboutissent souvent à des résultats mitigés sur les relations causales éducation/croissance et inégalité/croissance. Des analyses complémentaires se sont arrêtées longuement sur le concept du capital humain, ou encore celui de l'inégalité. En effet, le capital humain au sens large peut envelopper plusieurs facteurs immatériels « incorporés ». Il couvre non seulement les capacités intellectuelles de l'individu mais aussi celles physiques, mentales et sociales susceptibles d'influencer sa productivité. A leur tour, ces potentialités intellectuelles peuvent, en partie, être expliquées par l'éducation reçue. Cette dernière comprend aussi bien l'éducation formelle qui s'achève après la sortie du système éducatif que celle couvrant toutes les autres connaissances acquises, en dehors du système éducatif, tout au long de la vie.

De plus, si l'on retient uniquement la composante « éducation formelle » du capital humain dans la relation capital humain/croissance, on ne peut limiter les différentes facettes de cette éducation formelle au seul aspect quantitatif. En effet, celle-ci peut être développée dans au moins trois directions : l'augmentation du stock de l'éducation, l'amélioration de la qualité de l'enseignement et la réduction des inégalités éducatives.

Tout comme le « capital humain », « les inégalités » ont aussi un caractère multidimensionnel et englobent aussi bien l'aspect économique (salaires, revenus, épargne, patrimoine...), social (chômage, couverture sociale, services sanitaires...), culturel (éducation, lecture, loisirs...) que politique. Après la publication du rapport de la Banque mondiale de 2006 « Equity and Development » et l'introduction de l'IDHI⁹ par le PNUD en 2010, un intérêt grandissant pour l'analyse des inégalités « non monétaires » de façon globale et des inégalités éducatives en particulier a été constaté durant cette dernière décennie. Celles-ci sont considérées comme une variable aussi bien non quantitative de l'éducation que non monétaire de l'inégalité.

L'objectif de cette thèse, essentiellement empirique, est donc double : (i) retracer de manière stylisée l'évolution des inégalités éducatives sur une longue période et (ii) apporter un nouvel éclairage sur les deux relations précédemment citées à travers l'analyse d'un lien relativement peu exploré qui est celui entre les inégalités éducatives et le développement.

- ***Apports analytiques de la thèse***

Cette thèse, qui puise ses références analytiques dans les relations éducation/croissance et inégalité/croissance, questionne la pertinence du recours à la distribution de l'éducation pour mieux rendre compte de l'impact escompté du capital humain sur le développement économique. Nous étudions ainsi les conditions dans lesquelles la réduction de l'inégalité dans l'éducation peut être un moyen pour renforcer la place du capital humain dans les stratégies de croissance des pays en développement.

Nous allons attacher un intérêt spécial pour le caractère *non linéaire* de la relation entre les inégalités éducatives et le développement. Ainsi nous nous intéressons plus particulièrement à déceler l'existence d'éventuels seuils d'inégalité éducative à partir desquels s'entament de nouvelles phases de développement.

Cette thèse s'interroge également sur les arbitrages éducatifs liés à la distribution du capital humain. Etant donné que les ressources financières sont limitées, les politiques publiques doivent opérer des choix qui sont jugés plus bénéfiques pour le développement. Ainsi, est-il plus bénéfique pour le développement d'agir sur la moitié inférieure ou supérieure de la distribution du capital humain ? Autrement dit, faut-il encourager un modèle favorisant l'éducation de base à travers une plus grande massification ou plutôt un modèle élitiste insistant plus sur l'extension de l'enseignement supérieur ? Qu'est ce qui importe le plus pour le développement, est-ce la réduction des inégalités intragénérationnelles de l'éducation ou plutôt celle des inégalités intergénérationnelles ?

⁹ L'IDHI est l'indice de développement humain ajusté aux inégalités.

Cette thèse discute ainsi de nouvelles pistes pour orienter les politiques publiques en termes d'arbitrage éducatif dans les pays en développement.

Par ailleurs, nous identifions les différentes phases que traverse cette relation pour un large panel de pays en développement (PVD). Une évaluation plus fine au niveau de chacune des régions du monde est réalisée à travers notamment l'identification des seuils à partir desquels il semble apparaître un effet différent des inégalités éducatives sur la croissance. Parallèlement, et à titre de comparaison, certains résultats pour les pays développés seront aussi présentés.

Enfin, les constats relevés au niveau mondial pourraient en partie être influencés par l'hétérogénéité du panel transnational. Une analyse de la relation inégalités dans l'éducation / développement dans le cas d'un seul pays s'avère aussi pertinente à mener sur le plan analytique. Nous envisageons, dans le cadre de cette thèse, d'explorer cette relation dans le cas du Maroc en couvrant les trois niveaux de territorialité : régional, provincial et communal. Les questionnements soulevés au niveau transnational seront alors abordés au niveau infranational. De plus, un intérêt particulier sera accordé en parallèle à la relation inégalités dans l'éducation/développement humain étant donné son importance au niveau local.

- ***Apports méthodologiques de la thèse***

Comme on l'a précédemment souligné, les travaux sur le lien entre les inégalités d'éducation et le développement économique peuvent être distingués en deux types. Les travaux du premier type se réfèrent à la relation éducation/croissance (Lopez et al., 1998; Thomas et al., 2001; Park, 2006; Bowman, 2007; Sauer et Zagler, 2014) et retiennent aussi bien comme variables explicatives le niveau moyen d'éducation que l'inégalité dans l'éducation. Les travaux du second type, qui renvoient plutôt à la relation inégalité/croissance, introduisent eux aussi le niveau moyen d'éducation, dans la plupart des cas, avec les deux variables d'inégalité : celle de l'éducation et celle du revenu (Castelló et Doménech, 2002; Digdowiseiso, 2009; Rodríguez-Pose et Tselios, 2010; Simões et Duarte, 2010; Castelló, 2010a; Castelló, 2010b; Ilon, 2011; Földvári et Van Leeuwen, 2011).

Nous constatons ainsi que, pour ces deux types de travaux, les variables du niveau moyen et de l'inégalité d'éducation sont toutes les deux simultanément introduites dans la quasi-totalité des estimations effectuées sur la relation. Berthélemy et Arestoff (2003) signalent que ces estimations sont entachées d'un problème dû à la forte corrélation entre ces deux variables puisqu'une plus faible moyenne de l'éducation va de pair avec une plus grande inégalité de celle-ci. Pour contourner ce problème de corrélation entre ces deux variables, Gille (2015) a utilisé une méthodologie proposée par Berthélemy (2006) qui consiste à séparer la composante propre au niveau moyen d'éducation de la composante relative à l'inégalité de l'éducation. Gille (2015) signale que, même après correction, l'indice corrigé de l'inégalité en éducation demeure sensiblement corrélé au niveau moyen.

Nous proposons ainsi d'affiner les hypothèses sous-jacentes à la méthodologie de Berthélemy (2006). Dans ce sens, un cas plus général est présenté pour tester la persistance des résultats relevés par Berthélemy. En effet, l'auteur suppose une représentation simplifiée en quatre niveaux retenus (néant, primaire,

secondaire et supérieur). Celle-ci ne tient pas compte de l'inégalité au sein de chaque cycle d'enseignement. De plus, la durée d'études au sein de chaque cycle d'enseignement est présumée être la même pour les trois cycles.

Par ailleurs, la méthode de mesure de l'inégalité dans l'éducation que nous proposons apporte des éléments de réponse permettant d'éclairer la relation entre inégalités d'éducation/développement. Nous nous interrogeons également, dans le cadre de cette thèse, sur les hypothèses qui sous-tendent les différentes bases existantes sur les inégalités d'éducation et étudions leurs limites. Nous proposons à cet égard une nouvelle base mondiale qui présente, pour la première fois, plusieurs indices d'inégalité d'éducation par groupes d'âge pour un large panel mondial. Celle-ci tient compte de l'inégalité intra-cycles d'enseignement, de la variation et l'hétérogénéité dans les durées des cycles d'enseignement au cours du temps et entre les différents groupes d'âge et ce, pour l'ensemble du panel mondial.

Nous envisageons aussi de réexaminer l'hypothèse de linéarité présumée par la majorité des travaux. En effet, la relation entre les inégalités éducatives et la croissance économique a été le plus souvent testée en utilisant des modélisations paramétriques classiques où le lien est supposé linéaire. Nous analysons cette relation à travers des modèles non-paramétriques et semi-paramétriques en données de panel. Cette démarche permet d'ajuster au mieux le nuage de points de la relation tout en bénéficiant à la fois de l'intérêt des données de panel et des fonctions non paramétriques lisses et souples qui n'exigent pas de formes fonctionnelles prédéfinies.

4. Présentation du plan de la thèse

Cette thèse est structurée autour de trois chapitres. Le chapitre I présente une nouvelle base mondiale sur les inégalités d'éducation. Le chapitre II explore, à travers des spécifications non-paramétrique et semi-paramétrique, l'effet des inégalités d'éducation sur le développement économique. Le chapitre III est une application au cas du Maroc et présente une analyse de la relation au niveau infranational.

Le premier chapitre de cette thèse s'intéresse à la mesure des inégalités dans l'éducation. Il explore le cadre méthodologique pour la construction des indices d'inégalité dans l'éducation. Avant de proposer la nouvelle base de données sur les inégalités d'éducation, nous étudions les contraintes qui entravent l'élaboration de telles bases au niveau mondial. Une revue critique des différentes bases élaborées jusqu'à présent est aussi présentée afin de situer l'apport de notre nouvelle base mondiale.

Nous analysons aussi dans le cadre de ce chapitre le lien entre les indices d'inégalités d'éducation et le taux d'analphabétisme ainsi que celui entre l'indice de Gini de l'éducation et le niveau moyen de l'éducation. A cet égard, la formulation proposée par Berthélemy (2006) sera généralisée dans le cas de plusieurs niveaux d'éducation. Par la suite, nous déduisons, pour la première fois, le domaine de variation possible de l'indice de Gini de l'éducation selon la moyenne d'années de scolarisation, afin de déceler de manière générale la nature de la relation entre ces deux variables.

En se basant sur notre propre base de données, nous présentons à la fin de ce premier chapitre les traits marquants de l'évolution des inégalités éducatives entre 1950 et 2010. Les résultats seront détaillés par niveau de développement, par groupe d'âge et par zone géographique.

Le deuxième chapitre de cette thèse cherche à étudier la non linéarité de la relation entre les inégalités éducatives, mesurées selon plusieurs indices (les indices de Gini, de Theil, d'entropie généralisée d'ordre 0.5 et d'Atkinson d'ordre 1), et le développement économique pour un large panel de PVD entre 1955 et 2010. Après avoir dressé un état des lieux de la littérature théorique et empirique de cette relation, nous commençons par explorer l'existence de liens paramétriques non linéaires. Ensuite, nous proposons des spécifications non linéaires plus robustes. Une modélisation non-paramétrique et semi-paramétrique en données de panel sera ainsi menée dans le cadre de cette relation. Celle-ci nous permettra d'identifier, à partir des changements de concavité dans la trajectoire, les différentes phases que traverse le développement depuis le niveau le plus inégalitaire d'éducation jusqu'au niveau le moins inégalitaire observé. Nous nous attachons à repérer les traits communs et les dissemblances entre les régions du monde selon les seuils d'inégalité éducative et/ou de développement délimitant ces phases.

La dernière section de ce chapitre s'intéresse aux questions d'arbitrage éducatif et propose une démarche qui consiste à estimer des surfaces non-paramétriques basées sur des décompositions *within* et *between* de l'inégalité d'éducation globale. Cette démarche permet de déceler, dans un cadre non linéaire, les types d'inégalités d'éducation qui impactent le plus le développement économique. A cet effet, deux types de décomposition seront considérés : l'inégalité inter et intra moins éduqués / plus éduqués et l'inégalité intergénérationnelle et intragénérationnelle.

Au-delà du schéma général tracé sur le plan transnational dans les chapitres I et II, une étude d'un pays est menée au chapitre III. Il s'agit du Maroc, pour lequel nous disposons de données fines au plan régional, provincial et communal, et où aucune étude n'a traité les effets des inégalités éducatives sur le développement territorial.

Nous explorons, dans ce troisième chapitre, la non-linéarité de la relation en utilisant deux types de mesure du développement. Il s'agit du PIB par habitant et de l'indice de développement humain (IDH). Nous nous attacherons, dans un premier temps, à estimer au niveau régional l'indice de Gini de l'éducation et l'indice de développement humain et à en dégager les principaux faits stylisés. Dans un deuxième temps, nous menons deux types d'analyses complémentaires pour appréhender les effets des inégalités éducatives sur le développement territorial au Maroc : la première analyse concerne le niveau communal tandis que la seconde s'intéresse plutôt au niveau provincial.

Au niveau communal, nous nous attacherons à étudier l'impact des inégalités d'éducation sur le développement humain. A cet effet, nous proposons une base de données originale de 1534 communes marocaines en 2004 comportant à la fois l'indice de Gini de l'éducation, l'indice de développement humain ainsi que d'autres variables telles que les infrastructures et la croissance de la population.

Les estimations sont basées sur des données très désagrégées issues du recensement de la population marocaine de 2004.

La compilation de cette base de données permet de mettre en évidence l'ampleur des inégalités éducatives et des disparités de développement humain à un échelon très fin du territoire. Elle permet aussi de mener une analyse exploratoire des données spatiales (ESDA) afin de déceler l'existence d'une autocorrélation spatiale au Maroc sur des données communales. Finalement, nous explorons à travers un modèle économétrique spatial la nature de la relation entre inégalités éducatives et développement local.

Dans l'analyse menée au niveau provincial (61 provinces marocaines durant la période 1998-2013), nous adoptons la même démarche proposée dans le chapitre II. Il s'agira ainsi d'analyser les effets des inégalités d'éducation sur le développement économique par le biais d'une modélisation non-paramétrique et semi-paramétrique en données de panel.

Tout comme le chapitre II, nos estimations vont se baser sur les quatre indices d'inégalité d'éducation retenus (les indices de Gini, de Theil, d'entropie généralisée d'ordre 0.5 et d'Atkinson d'ordre 1). Ces derniers sont tous calculés sur la base de données très désagrégées issues des recensements de la population marocaine (1994 et 2004) et de l'ENDPR 2009-2010. Ainsi, nous construisons une base de données provinciales, contenant les estimations des quatre indices d'inégalité ainsi que leurs composantes *within* et *between* pour les deux types de décomposition proposés au chapitre II (au total 20 indices/composantes d'inégalité sont calculés).

Nous testons d'abord l'existence d'une tendance non linéaire par une modélisation polynomiale, non-paramétrique et semi-paramétrique. Nous identifions ensuite les différents types d'inégalités qui impactent le développement local au Maroc.

Chapitre I

La mesure des inégalités éducatives

- 1.1. Introduction
- 1.2. Comment mesurer les inégalités dans l'éducation ?
- 1.3. Proposition d'une nouvelle base mondiale sur les inégalités éducatives
- 1.4. Relations entre l'indice de Gini de l'éducation, le taux d'analphabétisme et la moyenne d'années de scolarisation
- 1.5. Tendances des inégalités éducatives selon les niveaux de développement des pays
- 1.6. Conclusion

Chapitre I

La mesure des inégalités éducatives¹⁰

1.1. Introduction

Les indices d'inégalité de revenu, largement utilisés dans la littérature, renseignent globalement sur le degré de concentration de la répartition des ressources monétaires perçues par les individus d'une population. Cependant, les revenus ne permettent pas, à eux seuls, de caractériser les inégalités de bien-être. Les indices d'inégalité dans la distribution de la santé ou de l'éducation, récemment introduits par le PNUD dans l'indice de développement humain ajusté aux inégalités (PNUD, 2010), permettent de mieux rendre compte en introduisant d'autres aspects de la nature multidimensionnelle des inégalités.

L'expansion de l'éducation dans un certain nombre de pays en développement, mesurée par les indices ordinaires de scolarisation, ne doit pas occulter le problème de la répartition inégale des niveaux d'instruction entre les individus. Le taux de scolarisation est un indicateur qui renseigne globalement sur le degré d'accès à l'éducation mais sous-estime les niveaux cumulés de scolarité. Le taux d'alphabétisation des jeunes, quant à lui, ne saisit que les premiers stades de l'accumulation du capital humain et ignore de ce fait les connaissances et les compétences acquises au-delà de l'éducation de base. La moyenne d'années de scolarisation, qui est fréquemment utilisée pour mesurer le niveau de capital humain, ne permet cependant pas de caractériser le niveau de concentration de la distribution d'éducation au sein d'une population.

Quoiqu'elles aient été largement étudiées dans le champ de la sociologie de l'éducation, les inégalités éducatives sont peu explorées dans le champ de l'économie de l'éducation. Au départ, comme le souligne Dessus (2001), le manque de bases de données sur les inégalités d'éducation comparables au niveau international est la principale raison. "The measure of dispersion of qualifications is more delicate to build. Ideally the good measure would be a Gini index of education, for which the years of education of each individual would replace the traditional measure of income. To our best knowledge, this measure is unfortunately *unavailable for international comparisons*" (Dessus, 2001).

La construction de bases internationales sur les inégalités éducatives a, depuis, fourni un support permettant l'émergence de nouveaux travaux. Une littérature très récente s'est développée pour explorer d'éventuelles liaisons causales entre les inégalités éducatives et diverses variables d'intérêt, notamment l'expansion de la scolarisation (Lim et Tang, 2008; Morrisson et Murin, 2010/2013; Meschi et Scervini, 2014), les inégalités de revenu (Checchi, 2004; Földvári et Van Leeuwen, 2014), la croissance économique (Castelló et Doménech, 2002; Castelló, 2010), la santé (Galea et al., 2007; Lê et al., 2010; Hori, 2011), la corruption (Patrawart, 2010), la cohésion sociale (Green et al, 2006; Vert, 2011), la migration (McKenzie et Rapoport, 2007), la démocratie (Castelló, 2008), la

¹⁰ Ce chapitre s'inspire d'un travail présenté dans Benaabdelaali et al. (2012)

volatilité macroéconomique (Checchi et García-Peñalosa, 2004), voire même la rébellion populaire (Bartusevičius, 2014).

Nous présentons en section 2 de ce chapitre une revue critique des différentes bases mondiales élaborées jusqu'à présent pour estimer les inégalités dans l'éducation. Dans la section 3, nous décrivons l'approche méthodologique que nous adoptons pour construire une nouvelle base de données des inégalités éducatives à l'échelle mondiale. Dans la section 4, nous proposons une formulation plus générale du lien entre l'indice de Gini, le niveau moyen de l'éducation et la durée cumulée de scolarisation. Dans la section 5 nous donnons un aperçu des tendances globales des inégalités éducatives à l'échelle mondiale.

1.2. Comment mesurer les inégalités dans l'éducation ?

1.2.1. La construction des indices d'inégalités éducatives : cadre méthodologique

De manière générale la construction d'un indice d'inégalité nécessite (i) de cerner la population d'intérêt sur laquelle l'analyse porte, (ii) de préciser sur quelle variable d'éducation l'inégalité est mesurée, et (iii) d'identifier l'entité sur laquelle l'observation de cette variable s'opère. Le premier critère concerne la population d'intérêt qui peut être, par exemple, l'ensemble des élèves de la sixième année de primaire de la province de Rabat en 2014, la population marocaine âgée de 15 ans et plus en 2004, ou bien, l'ensemble des élèves et étudiants en cours de scolarisation au Maroc en 2010.

Concernant le deuxième critère, on définit la variable d'éducation choisie pour évaluer l'inégalité. Le choix de cette dernière revêt une certaine importance, car elle détermine de manière systématique sur quoi l'inégalité est quantifiée. Ainsi l'inégalité peut être évaluée, par exemple, à partir du taux de scolarisation, du ratio élèves par classe, du taux de rétention, du ratio élèves par enseignant, du score du test en mathématiques, du niveau des dépenses en éducation dans l'enseignement scolaire, du volume horaire de scolarisation, du taux d'abandon ou du niveau maximum atteint d'éducation (*educational attainment*).

En troisième lieu, il s'agit de préciser l'entité d'observation appelée aussi entité d'analyse (troisième critère). Elle définit implicitement entre qui et qui l'inégalité est appréciée. Elle peut l'être entre ménages, individus, élèves, enseignants, classes, établissements, cycles d'enseignement, communes, provinces... Le manque d'information relatif à certains niveaux d'observation contraint de passer à un niveau supérieur à travers une agrégation (par exemple passer du niveau élève au niveau classe ou établissement). Le choix de l'entité d'observation est susceptible, toutefois, de modifier les résultats obtenus, même s'il s'agit à la fois de la même population d'intérêt et de la même variable d'éducation.

Plus encore, même s'il s'agit à la fois de la même population d'intérêt et de la même variable d'éducation et de la même entité d'observation mais avec une pondération différente les résultats changent. En effet, il faut signaler que le choix des poids revêt une importance essentielle dans le calcul des indices d'inégalité. Il convient de noter à ce stade que certains auteurs (Monchar, 1981) considèrent leurs observations comme étant équipondérées alors que d'autres utilisent des poids plus ou moins adéquats.

Sheret (1984) a repris les calculs des indices d'inégalités éducatives sur la base des mêmes taux de scolarisation utilisées par Monchar (1981) tout en tenant compte du poids de chacune des régions australiennes. Il a remarqué que les inégalités interrégionales en Australie entre 1958 et 1974 semblent montrer une tendance différente de celle mise en évidence par Monchar (1981) qui a considéré des régions équipondérées. Sheret (1984) a montré que ces inégalités interrégionales présentent plutôt une tendance à la hausse qu'elles soient mesurées par l'indice de Gini ou par le coefficient de variation.

Certes, les travaux sur les inégalités d'éducation qui ne prennent pas en compte le poids de chaque entité d'observation aboutissent à des résultats biaisés. Néanmoins, même les travaux ayant utilisé des pondérations n'ont pas forcément considéré les poids adéquats. Dans certains cas, le choix pertinent des bons poids revêt une certaine difficulté. Pour remédier à cette difficulté, il faut d'abord revenir à la formule de la moyenne de la variable d'éducation puis s'intéresser à l'indice d'inégalité de la variable en question. Nous présentons dans ce qui suit quatre exemples illustratifs.

Dans le premier exemple on prend comme variable d'éducation x "le niveau espéré de fin d'études". Les entités d'observation " i " sont donc des niveaux d'éducation. Ainsi l'observation $x_i^{Edu} = i$ tel que $i = 0..N$ ou N est le dernier niveau scolaire offert par le système éducatif en question. On note T_i le taux d'accès au niveau scolaire " i ". La moyenne de cette variable correspond donc à l'espérance de vie scolaire qui est égale à la somme des taux d'accès aux différentes années d'études. Pour trouver les pondérations adéquates, il faut reformuler la moyenne de la variable d'éducation de telle sorte à faire apparaître les observations x_i^{Edu} . Donc l'espérance de vie scolaire peut être reformulée comme suit :

$$\mu_X^{Edu} = \sum_{i=1}^N T_i = \sum_{i=1}^N T_i [i - (i - 1)] = \sum_{i=0}^N (T_i) i - \sum_{i=0}^N (T_{i+1}) i = \sum_{i=0}^N (T_i - T_{i+1}) i = \sum_{i=0}^N w_i^X X_i^{Edu}$$

Ainsi le poids du niveau scolaire " i " correspond à la différence entre les taux d'accès de deux niveaux scolaires successifs $w_i^X = (T_i - T_{i+1})$, puisque la somme $\sum_{i=0}^N (T_i - T_{i+1}) = T_0 - T_{N+1} = 1$.

Dans le deuxième exemple la variable d'éducation x concerne "le ratio élèves par enseignant". Rao et Jani (2011) ont utilisé l'indice de Gini des ratios élèves par enseignants pour évaluer les inégalités d'éducation inter-États indonésiens. La pondération utilisée par les auteurs est le poids des élèves (rapport des élèves appartenant à chaque État sur l'ensemble des élèves indonésiens). Or ce poids n'est pas adéquat. En effet, dans ce cas, les entités d'observation " i " sont les États indonésiens, la population d'intérêt est l'Indonésie, l'observation x_i^{Edu} est le ratio élèves par enseignants à l'État " i " qui est égal au rapport de l'effectif des élèves de l'État " i " sur l'effectif des enseignants de l'État " i ". Dans ce cas $x_i^{Edu} = \frac{E_i}{P_i}$ où E_i représente l'effectif des élèves appartenant à l'État " i " et P_i représente l'effectif des enseignants appartenant au même État " i ". La moyenne nationale n'est autre que la somme de tous les élèves rapportée à la somme de tous les enseignants. Ainsi on peut retrouver les poids à partir de la formule de la moyenne comme suit :

$$\mu_X^{Edu} = \frac{\sum_{i \in \text{Régions}} E_i}{\sum_{j \in \text{Régions}} P_j} = \sum_{i \in \text{Régions}} \left[\frac{P_i}{\sum_{j \in \text{Régions}} P_j} \right] \left(\frac{E_i}{P_i} \right) = \sum_{i \in \text{Régions}} w_i^X X_i^{Edu}$$

D'où $w_i^X = \left[\frac{P_i}{\sum_{j \in \text{Régions}} P_j} \right]$. Ainsi les auteurs devraient plutôt pondérer les observations par le poids des enseignants (rapport des enseignants appartenant à chaque État sur l'ensemble des enseignants de l'Indonésie) et non par le poids des élèves.

Dans le troisième exemple, la variable d'intérêt x est le taux brut de scolarisation, la population d'intérêt est la population des scolarisés, les entités d'observation " i " sont les cycles d'enseignement $\{i=1$ pour le cycle primaire, $i=2$ pour le cycle secondaire collégial, $i=3$ pour le cycle secondaire qualifiant et $i=4$ pour le cycle supérieur}, l'observation X_i^{Edu} $i=1..4$ représente le taux brut de scolarisation au cycle " i ". De ce fait, $X_i^{Edu} = \frac{SCOL_i}{Pop_{a(i)}}$ où $SCOL_i$ représente l'effectif des scolarisés au cycle " i " et $Pop_{a(i)}$ représente la population du groupe d'âge $a(i)$ correspondant officiellement au cycle d'enseignement " i ".

La moyenne de cette variable d'éducation μ_X^{Edu} n'est autre que le taux brut combiné de scolarisation¹¹. Ainsi on peut retrouver les poids à partir de la formule de la moyenne :

$$\mu_X^{Edu} = \frac{\sum_{i \in 1..4} SCOL_i}{\sum_{j \in 1..4} Pop_{a(j)}} = \sum_{i \in 1..4} \left[\frac{Pop_{a(i)}}{\sum_{j \in 1..4} Pop_{a(j)}} \right] \left(\frac{SCOL_i}{Pop_{a(i)}} \right) = \sum_{i \in 1..4} w_i^X X_i^{Edu}$$

D'où $w_i^X = \left[\frac{Pop_{a(i)}}{\sum_{j \in 1..4} Pop_{a(j)}} \right]$ est le poids du cycle " i " qui correspond à la population du groupe d'âge correspondant officiellement au cycle d'enseignement " i " rapportée à la population de l'ensemble des groupes d'âge correspondant aux différents cycles d'enseignement (primaire, secondaire collégial, secondaire qualifiant et supérieur).

Prenons aussi comme dernier exemple un cas qui fera l'objet d'une illustration¹² dans la deuxième section et qui montre l'ampleur du biais qui peut être engendré suite à la non prise en compte d'une pondération adéquate. Il nous apparaît que Barro et Lee (2013) utilisent une pondération non adéquate pour agréger les durées cumulées de scolarisation par rapport aux différentes tranches d'âge. Ainsi, étant donné le couple (c, t) qui représente un pays " c " à une date " t ", la variable d'éducation en question est le nombre nécessaire d'années cumulées de scolarisation pour atteindre un niveau d'éducation " j " par la tranche d'âge " i " noté $X_{i,j}^{Edu}$. La population d'intérêt est la population âgée de 15 ans et plus et les entités d'observation " i " sont les tranches d'âge $\{i=1$ pour la tranche d'âge 15-19 ans, ..., $i=13$ pour la tranche d'âge 75 ans et plus}. On note $p_{i,j}$ (p_j^{15+}) la proportion de la population de la tranche d'âge " i " (15 ans et plus) ayant atteint le niveau d'éducation " j " tel que $\sum_{i \in 1..13} p_{i,j} = 1$ ($\sum_{i \in 1..13} p_j^{15+} = 1$), l_i le poids de la population appartenant à la tranche d'âge " i " (Pop_i) dans la population âgée de 15 ans et plus (Pop^{15+}) tel que $l_i = Pop_i / Pop^{15+}$ et μ_X^{Edu} la moyenne de notre variable d'éducation $X_{i,j}^{Edu}$ par rapport aux tranches d'âge. L'objectif est d'identifier $w_{i,j}^X$ tel que $\mu_X^{Edu} = \sum_{i=1}^{13} w_{i,j}^X X_{i,j}^{Edu}$. Ainsi, comme $p_j^{15+} = \sum_{i=1}^{13} l_i p_{i,j}$, on pourrait a priori penser que la moyenne μ_X^{Edu} serait de même égale à $\sum_{i=1}^{13} l_i X_{i,j}^{Edu}$. Or malgré le fait que cette pondération ($w_{i,j}^X = l_i$) ait été utilisée dans l'ensemble des versions de la base de données de Barro et Lee (2013) $\{1.0, 1.1, 1.2, 1.3$ et $2.0\}$, elle ne semble pas la plus appropriée. En effet, en partant des formules de la moyenne d'années de

¹¹ Le taux brut combiné de scolarisation a déjà été utilisé par le PNUD dans l'ancienne méthodologie du calcul de la composante éducation de l'IDH (avant 2010).

¹² Il s'agit du cas de la Bulgarie en 1990 à partir des données de la base de Barro et Lee (2013).

scolarisation de la tranche d'âge "i" (MAS_i) et de celle de la population âgée de 15 ans et plus (MAS^{15+}), on retrouve les bons poids relatifs comme suit :

$$\begin{cases} MAS^{15+} = \sum_{i=1}^{13} l_i MAS_i & (1) \\ MAS_i = \sum_{j=1}^n p_{i,j} X_{i,j}^{Edu} & (2) \\ MAS^{15+} = \sum_{j=1}^n p_j^{15+} \mu_j^{Edu} & (3) \end{cases} \Rightarrow \begin{cases} MAS^{15+} = \sum_{j=1}^n p_j^{15+} \sum_{i=1}^{13} \frac{l_i p_{i,j}}{p_j^{15+}} X_{i,j}^{Edu} & (1) \text{ avec } (2) \\ MAS^{15+} = \sum_{j=1}^n p_j^{15+} \mu_j^{Edu} & (3) \end{cases}$$

$$\Rightarrow \mu_j^{Edu} = \sum_{i=1}^{13} \frac{l_i p_{i,j}}{p_j^{15+}} X_{i,j}^{Edu} \neq \sum_{i=1}^{13} l_i X_{i,j}^{Edu}$$

Donc $w_{i,j}^X = \left[\frac{l_i p_{i,j}}{\sum_{i=1}^{13} l_i p_{i,j}} \right]$ représente la pondération pertinente de la tranche d'âge "i" qui est différente de celle utilisée par de Barro et Lee (2013) $\{w_{i,j}^X \neq l_i\}$.

Une fois la population d'intérêt, la variable d'éducation, l'entité d'observation et son poids correspondant bien définis, différents indicateurs de mesure peuvent être calculés pour caractériser les inégalités d'éducation. On note ainsi X_i^{Edu} l'observation de la variable d'éducation X pour l'entité d'observation "i" ayant le poids w_i^X et appartenant à une population d'intérêt P . L'indice d'inégalité d'éducation I est défini comme une application continue qui associe à toute distribution d'une variable d'éducation X^{Edu} un nombre réel $I(X^{Edu})$ représentant le degré d'inégalité dans la distribution X^{Edu} . On note respectivement μ_X^{Edu} et ϱ_X^{Edu} les moyennes arithmétique et géométrique de X^{Edu} telles que : $\mu_X^{Edu} = \sum_{i \in P} w_i^X X_i^{Edu}$ et $\varrho_X^{Edu} = \prod_{i \in P} X_i^{Edu w_i^X}$. Le tableau n°1 présente ainsi les formules des indices les plus utilisés.

Tableau 1: Formules des indicateurs d'inégalité d'éducation les plus utilisés

L'indice d'inégalité d'éducation	Sigle	Formule
Rapports interquartiles	$RIQ_p(X^{Edu})$	$\frac{Q_X^{Edu}(1-p)}{Q_X^{Edu}(p)}$
Etendues ou écarts interquartiles	$EIQ_p(X^{Edu})$	$Q_X^{Edu}(1-p) - Q_X^{Edu}(p)$
Ecart-type	$\sigma(X^{Edu})$	$\sqrt{\sum_{i \in P} w_i^X (X_i^{Edu} - \mu_X^{Edu})^2}$
Coefficient de variation	$CV(X^{Edu})$	$\frac{1}{\mu_X^{Edu}} \sqrt{\sum_{i \in P} w_i^X (X_i^{Edu} - \mu_X^{Edu})^2}$
Ecart-type du log	$\sigma_{\ln}(X^{Edu})$	$\sqrt{\sum_{i \in P} w_i^X (\ln(X_i^{Edu}) - \ln(\varrho_X^{Edu}))^2}$
Indice de Gini	$G(X^{Edu})$	$\frac{1}{2\mu_X^{Edu}} \sum_{i \in P} \sum_{j \in P} w_i^X X_i^{Edu} - X_j^{Edu} w_j^X$
Déviation logarithmique moyenne	$GE_0(X^{Edu})$	$\sum_{i \in P} w_i^X \ln\left(\frac{\mu_X^{Edu}}{X_i^{Edu}}\right)$
Indice de Theil	$GE_1(X^{Edu})$	$\sum_{i \in P} w_i^X \frac{X_i^{Edu}}{\mu_X^{Edu}} \ln\left(\frac{X_i^{Edu}}{\mu_X^{Edu}}\right)$
Entropie généralisée	$GE_\alpha(X^{Edu})$	$\frac{1}{\alpha(\alpha-1)} \left[\left(\sum_{i \in P} w_i^X \left(\frac{X_i^{Edu}}{\mu_X^{Edu}} \right)^\alpha \right) - 1 \right]$
Atkinson d'ordre 1	$A_1(X^{Edu})$	$1 - \frac{\varrho_X^{Edu}}{\mu_X^{Edu}}$
Atkinson	$A_\varepsilon(X^{Edu})$	$1 - \left[\sum_{i \in P} w_i^X \left(\frac{X_i^{Edu}}{\mu_X^{Edu}} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}$

Source: Auteur.

Nous avons fait le choix de distinguer entre deux types d'indicateurs : le premier concerne ceux qu'on qualifie d'élémentaires et le second comporte les indices « purs » d'inégalités. Les indicateurs du premier type ne vérifient pas toutes les propriétés usuelles de l'approche axiomatique des indices d'inégalités et s'appuient généralement sur un simple rapport entre deux moyennes. Ces indicateurs concernent principalement les disparités entre catégories (genre, milieu, groupe ethnique, statut migratoire...) comme par exemple le rapport du taux de scolarisation en milieu rural sur celui en milieu urbain, le rapport de la moyenne d'années de scolarisation des migrants sur celle des natifs ou bien le rapport du taux d'achèvement des enfants d'ouvriers sur celui des cadres supérieurs.

Les indicateurs du second type sont considérés comme des mesures d'inégalités proprement dites (indices purs d'inégalités) qui vérifient les axiomes¹³ des indices d'inégalités et se basent sur la distribution de la variable d'éducation. Ces derniers comportent d'une part les indices absolus, dont l'unité de mesure est la même que celle de la variable d'éducation en question; d'autre part les indices relatifs, grandeurs sans dimension, qui varient généralement entre 0 et 1. Nous présentons ainsi dans ce qui suit une revue de littérature sur les indices d'inégalité d'éducation les plus utilisés.

1.2.2. Les indices d'inégalités éducatives : revue de littérature

Dans le tableau n°2, nous présentons quelques travaux sur les inégalités éducatives. Nous nous attacherons dans ce qui suit à préciser pour chacun des travaux, la variable d'éducation, l'entité d'observation, la population d'intérêt considérée ainsi que l'échantillon d'analyse¹⁴. De plus, il sera aussi question, pour certains indices, d'analyser leur conformité aux axiomes des indices d'inégalité ainsi que de leur comportements vis-à-vis des changements dans les queues de la distribution.

Ainsi, les rapports interquantiles $RIQ_p(X^{Edu})$ sont des indices relatifs qui peuvent être calculés en rapportant le niveau agrégé de la fraction p détenant les valeurs les plus élevées de la distribution X^{Edu} à celui de la fraction p détenant les valeurs les plus faibles de cette même distribution. De même, les étendues interquantiles appelées aussi écarts ou intervalles interquantiles sont des indices absolus. Ils sont calculés en faisant la différence entre le niveau agrégé de la fraction p détenant les valeurs les plus élevées de la distribution X^{Edu} et celui de la fraction p détenant les valeurs les plus faibles.

¹³ La normalisation, la symétrie, le principe de transferts et l'invariance à l'échelle ou l'invariance par translation.

¹⁴ L'expression "échantillon d'analyse" dans ce chapitre désigne la réunion des populations d'intérêt qui forme l'objet d'analyse dans chacun des travaux sur les inégalités éducatives.

Tableau 2: Quelques travaux sur les inégalités éducatives selon divers variables d'éducation, entités d'observation, indices d'inégalité et populations d'intérêt

Auteurs	Echantillon d'analyse	Les trois critères d'un indice d'inégalité d'éducation			Indices d'inégalité
		Variables d'éducation	Population d'intérêt	Entité d'observation	
			Inégalités intra	Inégalités inter	
López et al. (1998)	12 pays, 1970-1995	Le niveau d'éducation maximum atteint	Pays	Individus	<ul style="list-style-type: none"> • Le coefficient de variation • L'écart-type du log • L'indice de Gini
Betts et al. (2000)	Californie, 1998	<ul style="list-style-type: none"> • Le nombre d'années d'expérience des enseignants • Le plus haut niveau de diplôme des enseignants • La taille moyenne des classes • Les scores des élèves en mathématiques et en lecture 	État (Californie)	Classes	Le rapport et l'étendue interquartiles
Tansel et Güngör (2000)	65 provinces de la Turquie, 1980 et 1994	Le taux brut de scolarisation	Province	Cycles	La déviation logarithmique moyenne
Merle (2002)	France, 1988-1998	L'âge de fin d'étude	Pays (France)	Etudiants	Les rapports et étendues interdéciles
Zhang et Li (2002)	76 pays, 1960-1990	La moyenne d'années de scolarisation	Ensemble de pays	Pays	<ul style="list-style-type: none"> • Le coefficient de variation • L'indice de Gini • L'écart-type du log
Thomas et al. (2003)	La population âgée de 15 ans et plus pour 140 pays, 1960-2000	Le niveau d'éducation maximum atteint	Pays	Individus	<ul style="list-style-type: none"> • L'indice de Gini • L'indice de Theil • L'écart-type
Iatarola et Stiefel (2003)	New York	<ul style="list-style-type: none"> • Les dépenses de fonctionnement par élève • La totalité des dépenses par élève • Le ratio élèves par enseignant • Les dépenses moyennes sur les salaires des enseignants • Le pourcentage des enseignants certifiés pour enseigner 	État (New York)	Ecoles	Le coefficient de variation
Jamal et Khan (2005)	Pakistan, 1998	<ul style="list-style-type: none"> • Le taux spécifique de scolarisation au primaire • Le taux spécifique de scolarisation au secondaire • Le taux spécifique de scolarisation au supérieur • Le taux d'analphabétisme des 10 ans et plus 	Pays (Pakistan)	Provinces	<ul style="list-style-type: none"> • Le rapport du maximum sur le minimum • Le coefficient de variation • L'indice de Gini
Sahn et Younger (2007)	38 pays en 1999 & 49 pays en 2003	<ul style="list-style-type: none"> • Le score des élèves en mathématiques • Le score des élèves en sciences 	Pays	Elèves	<ul style="list-style-type: none"> • La déviation logarithmique moyenne • L'indice de Theil • L'entropie généralisée d'ordre 2
Sherman et Poirier (2007)	16 pays les plus peuplés au monde	<ul style="list-style-type: none"> • Taux de scolarisation primaire/secondaire • Les dépenses par élève au primaire & secondaire • Les ratios élèves par enseignants primaire/secondaire 	Pays	Régions	<ul style="list-style-type: none"> • Le coefficient de variation • L'indice de McLoone ajusté • L'indice de Gini
Collins (2009)	24 provinces, Cambodge 1998-2003	Le taux net de scolarisation au primaire, collégial, lycée	Province et cycle	Niveaux scolaires	L'indice de Gini
Rao et Jani (2011)	Malaisie, 1986-2006	Le ratio élèves par enseignant	Pays (Malaisie)	États	L'indice de Gini
Felouzis et al. (2011)	12 cantons suisses, 2003	Le score des élèves en mathématiques en 9 ^{ème} année	Canton	Elèves	Le coefficient de variation
Chow et Shen (2014)	Chine	les dépenses d'éducation par habitant	Pays (Chine)	Provinces	L'écart-type du log
Catin et al. (2015)	116 régions d'Afrique australe et de l'est, (15 pays), 2007	La moyenne du score en mathématiques et en lecture	Région	Elèves	<ul style="list-style-type: none"> • L'indice de Gini • L'indice de Theil • L'indice d'Atkinson

Source : Auteur.

Betts et al. (2000) ont utilisé le rapport et l'étendue interquartiles¹⁵ pour mesurer les inégalités entre les écoles en Californie en 1998 moyennant des variables de ressources éducatives et de résultats scolaires, notamment le nombre d'années d'expérience des enseignants, le plus haut niveau de diplôme des enseignants, la taille moyenne des classes ainsi que les scores des élèves en mathématiques et en lecture. De même, Merle (2002) a utilisé quatre rapports et intervalles interdéciles¹⁶ pour évaluer les inégalités entre les étudiants français entre 1988 et 1998 en termes d'allongement de la durée de scolarisation mesurée par l'âge de fin d'études.

Les rapports et les étendues interquartiles sont très utilisés puisqu'ils sont des indices faciles à calculer mais ne sont toutefois pas conformes à l'approche axiomatique des indices d'inégalité et ne reflètent pas l'information pour l'ensemble de la répartition. En effet, ils se basent uniquement sur une partie de la distribution X^{Edu} et ignorent de ce fait la tendance centrale.

Le coefficient de variation $CV(X^{Edu})$, est beaucoup plus connu dans les analyses de dispersion. Il a été largement utilisé dans les premiers travaux s'intéressant aux inégalités scolaires. Psacharopoulos (1977) et Ram (1982) l'ont calculé sur les effectifs scolarisés afin d'évaluer les inégalités inter-cycles (primaire, secondaire et supérieur) pour 49 pays en 1973 et 64 pays en 1977 respectivement. Pour mesurer les inégalités éducatives interrégionales en Australie, Monchar (1981) l'a employé pour le taux de scolarisation combiné des cycles primaire et secondaire. López et al. (1998) quant à eux l'ont utilisé pour le niveau d'éducation maximum atteint pour 12 pays entre 1970 et 1995. Iatarola et Stiefel (2003) l'ont utilisé pour apprécier les inégalités des ressources éducatives entre les écoles de New York. Cinq variables ont été retenues par ces auteurs : les dépenses de fonctionnement par élève, l'ensemble des dépenses par élève, le ratio élèves par enseignant, les dépenses moyennes sur les salaires des enseignants et le pourcentage des enseignants certifiés pour enseigner. Felouzis et al. (2011) l'ont utilisé pour mesurer les inégalités de compétences entre élèves dans douze cantons suisses en 2003.

Le coefficient de variation, continue tout de même à être utilisé puisqu'en plus d'être considéré comme un indicateur qui synthétise l'information pour l'ensemble de la répartition, il est conforme à l'approche axiomatique des indices d'inégalité et est facilement calculable en divisant l'écart-type par la moyenne. Toutefois, cet indicateur reste assez sensible à la présence de valeurs élevées et toute réduction de ces dernières implique une chute significative de ce coefficient.

L'écart-type du logarithme $\sigma_{\ln}(X^{Edu})$, qui est aussi un indicateur de mesure relative¹⁷, est aussi présent dans certains travaux sur les inégalités éducatives. Chow et Shen (2014) l'ont utilisé pour des données sur les dépenses d'éducation par habitant au niveau provincial dans le cas de la Chine en 2002. Ces auteurs ont ainsi

¹⁵ Le rapport et l'étendue interquartiles correspondent respectivement à $RIQ_{25\%}(X^{Edu})$ et à $EIQ_{25\%}(X^{Edu})$. Dans le cas où $p = 20\%$, $p = 10\%$, $p = 5\%$ ou $p = 1\%$, ces indices sont appelés respectivement rapport interquintiles, interdéciles, intervingtiles et intercentiles.

¹⁶ Pour les rapports interdéciles il a utilisé D9/D1, D8/D2, D7/D3, D6/D4 et pour les étendues interdéciles il a utilisé D9-D1, D8-D2, D7-D3, D6-D4.

¹⁷ Cet indicateur est une mesure relative (invariante à l'échelle) alors que l'écart-type d'une variable est communément reconnu comme étant une mesure absolue. En effet, comme la différence entre le logarithme de l'observation X_i^{Edu} et le logarithme de la moyenne géométrique φ_X^{Edu} (cf. Tableau n° 2) est équivalente au logarithme de leur rapport $\ln(X_i^{Edu}) - \ln(\varphi_X^{Edu}) = \ln(X_i^{Edu}/\varphi_X^{Edu})$, l'écart-type du logarithme $\sigma_{\ln}(X^{Edu})$ satisfait donc au principe d'invariance à l'échelle de mesure.

montré que les inégalités interprovinciales d'éducation dans l'enseignement supérieur sont plus importantes que les inégalités interprovinciales de revenu. De même, Zhang et Li (2002) l'ont aussi appliqué pour les moyennes d'années de scolarisation des pays du monde afin d'appréhender les inégalités inter-pays d'éducation. Toutefois, contrairement au coefficient de variation d'une variable d'éducation, l'écart-type de son logarithme tend à affecter plus de poids aux valeurs faibles puisque l'introduction du logarithme a pour effet de rapprocher les valeurs de l'extrémité supérieure de la distribution.

Si le coefficient de variation et l'écart-type du logarithme sont sensibles respectivement aux extrémités supérieure et inférieure de la distribution, l'indice de Gini quant à lui est peu sensible aux deux queues de la répartition de la variable d'éducation. En effet, il accorde plutôt plus de poids aux changements aux alentours du mode de la distribution (où la densité des individus est la plus élevée).

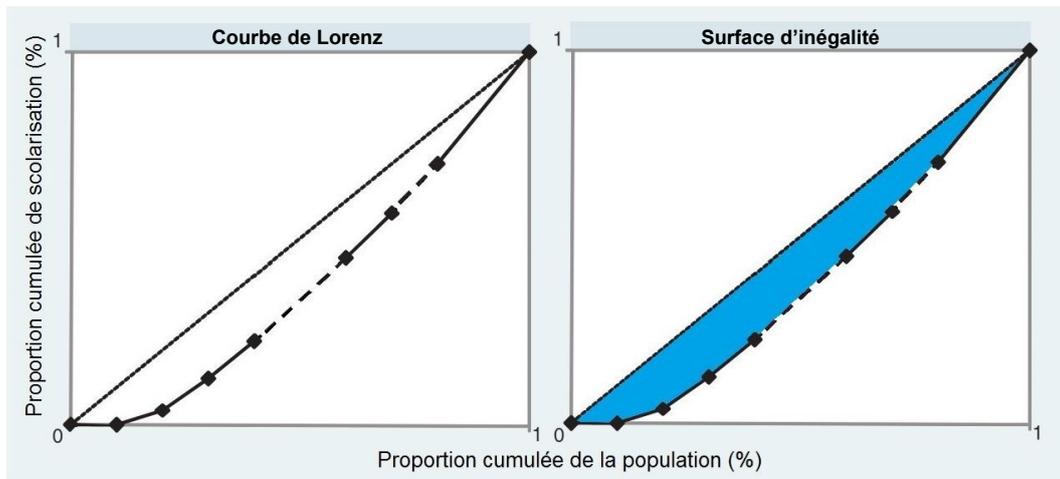
L'indice de Gini de l'éducation est la mesure la plus reconnue pour apprécier le niveau de concentration des inégalités éducatives au sein d'une population d'intérêt donnée. Il est égal à la moitié de la moyenne des écarts absolus entre les observations prises deux à deux rapportée à la moyenne des observations. Il peut aussi être défini en faisant référence à la courbe de Lorenz de l'éducation obtenue en reportant en ordonnée le pourcentage cumulé de la scolarisation et en abscisse le pourcentage cumulé de la population (courbe tracée à la gauche du graphique n° 2). L'indice de Gini de l'éducation est ainsi égal à deux fois la surface entre la courbe de Lorenz et la diagonale (surface colorée à la droite du graphique n°2).

Ainsi, Jamal et Khan (2005) l'ont utilisé pour analyser les inégalités interprovinciales au Pakistan en 1998 en considérant le taux d'analphabétisme de 10 ans et plus et les taux spécifiques de scolarisation du primaire, secondaire et du supérieur. Sherman et Poirier (2007) ont évalué les inégalités interrégionales scolaires des ressources financières et humaines dans les 16 pays les plus peuplés au monde en utilisant l'indice de Gini pour le ratio élèves par enseignant et les dépenses d'éducation par élève.

Collins (2009) a apprécié, pour chacune des 24 provinces et municipalités du Cambodge, les inégalités entre les niveaux scolaires dans chaque cycle d'enseignement en utilisant l'indice de Gini pour les taux net de scolarisation au primaire, collège et lycée. Rao et Jani (2011) ont estimé les inégalités entre les Etats indonésiens pour la période 1986-2008 à travers l'indice de Gini des ratios élèves-enseignants. Catin et al. (2015) ont calculé un indice de Gini de l'éducation pour évaluer les inégalités éducatives au sein de chacune des 116 régions de 15 pays de l'Afrique Australe et de l'Est sur la base de données des élèves en mathématiques et en lecture.

L'entropie généralisée permet aussi de rendre compte des inégalités d'éducation après avoir caractérisé, au travers d'un paramètre noté α , le degré de sa sensibilité aux changements en haut de la distribution. Ainsi, plus α est élevé, plus l'indice d'entropie $GE_{\alpha}(X^{Edu})$ est sensible aux changements à l'extrémité supérieure de la variable d'éducation. Ainsi, la déviation logarithmique moyenne appelée aussi l'écart moyen des logarithmes et l'indice de Theil (1967) appartiennent tous les deux à la famille des indices d'entropie généralisée et correspondent respectivement au cas où $\alpha = 0$ et $\alpha = 1$.

Graphique 2 : Courbe de Lorenz de l'éducation et l'interprétation géométrique de l'indice de Gini de l'éducation



Source : Auteur.

Tansel et Güngör (2000) ont utilisé la déviation logarithmique moyenne ($GE_0(X^{Edu})$) pour évaluer les inégalités inter-cycles des taux brut de scolarisation pour chacune des 65 provinces de la Turquie entre 1980 et 1994. Sahn et Younger (2007) ont utilisé les indices d'entropie généralisée d'ordre 0, 1 et 2 ($GE_0(X^{Edu})$, $GE_1(X^{Edu})$, $GE_2(X^{Edu})$) pour analyser les inégalités entre les scores des élèves en mathématiques et en sciences au sein des pays participant aux enquêtes internationales d'évaluation des acquis des élèves TIMSS de 1999 et 2003.

De même, tout comme la famille des indices d'entropie généralisée, les indices d'Atkinson offrent la possibilité de paramétrer l'aversion à l'inégalité et permettent de caractériser la sensibilité à l'extrémité inférieure de la variable d'éducation. Ainsi, plus la valeur choisie pour ϵ est élevée, plus l'indice d'Atkinson est sensible aux valeurs faibles. On considère généralement pour ϵ des valeurs inférieures ou égales à 2.

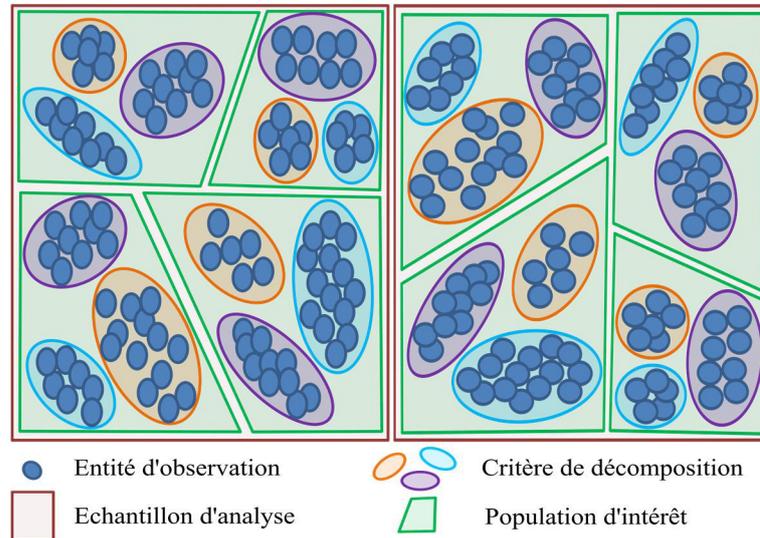
Certains auteurs ont utilisé la famille des indices d'Atkinson pour évaluer les inégalités éducatives. En utilisant la variable "niveau d'éducation maximum atteint", le PNUD utilise depuis 2010 $A_1(X^{Edu})$ pour quantifier la perte due aux inégalités éducatives au sein de 152 pays. Fidalgo et al. (2010) ont calculé pour la même variable un indice d'Atkinson d'ordre 0.5 ($A_{0.5}(X^{Edu})$) dans 20 régions portugaises entre 1986 et 2005. Meschi et Scervini (2014) ont estimé aussi pour la même variable les indices d'Atkinson d'éducation d'ordre 0.5, 1 et 2 ($A_{0.5}(X^{Edu})$, $A_1(X^{Edu})$, $A_2(X^{Edu})$).

1.2.3. Décomposition des indices d'inégalités éducatives

Au-delà de la mesure de l'inégalité, certains travaux se sont intéressés à décomposer les inégalités d'éducation selon un critère donné afin d'évaluer la contribution des inégalités inter- et intra-groupes à l'inégalité totale constatée. La figure n° 1 illustre le schéma de différenciation des entités d'observation selon un critère de décomposition de trois groupes. Ce dernier, qui peut être par exemple le milieu, le genre, les tranches d'âge, les provinces, le niveau de pauvreté, de revenu ou de développement..., partitionne la population d'intérêt en groupes distincts. Ainsi le nombre de groupes dépend du choix de ce critère, pour le cas par exemple du milieu ou du genre la population est scindée en deux groupes. Bien évidemment

on ne peut procéder à une décomposition de l'inégalité d'éducation que si la variable d'éducation est aussi bien renseignée dans les groupes définis par ce critère que dans les entités d'observations appartenant à chacun de ces groupes.

Figure 1 : Schéma de différenciation des entités d'observation selon un critère donné de décomposition (cas d'un critère de trois groupes)



Source : Auteur.

Le tableau n° 3 présente quelques travaux sur les décompositions des inégalités d'éducation. Certains travaux, comme Zhang et Li (2002), Mesa (2007), Qian et Smyth (2008), Gustave et Joubert (2012), Zamora et Dorado (2015), ont décomposé l'indice de Gini de l'éducation en deux composantes (*within* et *between*). En effet, après avoir calculé la contribution des inégalités intragroupes (*within*) les auteurs déduisent la contribution des inégalités intergroupes par soustraction de la composante *within* à l'inégalité totale. D'autres travaux comme Foldvari et van Leeuwen (2010), Agrawal (2014) et Yang et al. (2014) ont décomposé l'indice de Gini de l'éducation en trois composantes. Ainsi, les auteurs calculent les contributions des inégalités intragroupes (*within*) et intergroupes (*between*), puis déduisent la contribution du troisième élément de la décomposition par soustraction des deux composantes *within* et *between* à l'inégalité totale. Ce dernier, qui informe sur l'importance du chevauchement entre les distributions de chacun des groupes, est appelé "la contribution des inégalités d'éducation inhérentes à l'intensité de la transvariation".

D'autres auteurs ont utilisé la famille des indices d'entropie généralisée pour décomposer les inégalités éducatives. Ainsi, Sahn et Stifel (2003), Corcoran et Evans (2007), Lin et Yang (2009), Adroque (2013), Umar et al. (2014a) et Tan et al. (2015) ont utilisé l'indice de Theil de l'éducation pour décomposer les inégalités d'éducation respectivement selon le milieu, les États, les tranches d'âge, les provinces, le milieu et les écoles. Borooh et Knox (2015) ont décomposé la déviation logarithmique moyenne ($GE_0(X^{Edu})$) selon le type des écoles irlandaises tandis que Sahn et Younger (2007) ont décomposé trois indices d'entropie généralisée pour caractériser les inégalités intra et inter pays des acquis scolaires (tableau n° 3).

Nous présentons ainsi les formules de décomposition de quelques indices d'inégalité :

$$\begin{cases}
Gini = Gini_W + Gini_B + Gini_T \\
Gini_W = \left[\sum_{k=1}^K \frac{q_k^2 \mu_k}{\mu} Gini_k \right] \\
Gini_B = \left[\frac{1}{2\mu} \sum_{k=1}^K \sum_{k'=1}^K q_k |\mu_k - \mu_{k'}| q_{k'} \right] \\
GE(\alpha) = GE(\alpha)_W + GE(\alpha)_B \\
GE(\alpha)_W = \left[\sum_{k=1}^K q_k \left(\frac{\mu_k}{\mu} \right)^\alpha GE_k(\alpha) \right] \\
GE(\alpha)_B = \begin{cases} \left[\frac{1}{\alpha(1-\alpha)} \sum_{k=1}^K q_k \left(\frac{\mu_k}{\mu} \right)^\alpha - 1 \right] & \text{si } \alpha \neq 0 \text{ et } 1 \\ \left[\sum_{k=1}^K q_k \ln \left(\frac{\mu}{\mu_k} \right) \right] & \text{si } \alpha = 0 \\ \left[\sum_{k=1}^K \left(\frac{q_k \mu_k}{\mu} \right) \ln \left(\frac{\mu_k}{\mu} \right) \right] & \text{si } \alpha = 1 \end{cases} \\
A(\varepsilon) = A(\varepsilon)_W + A(\varepsilon)_B - A(\varepsilon)_W \times A(\varepsilon)_B \Leftrightarrow [1 - A(\varepsilon)] = [1 - A(\varepsilon)_W] \times [1 - A(\varepsilon)_B] \\
A(\varepsilon)_W = \begin{cases} \left[1 - \frac{1}{\left[\sum_{k=1}^K q_k \left(\frac{\mu_k}{\mu} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}} \sum_{k=1}^K q_k \left(\frac{\mu_k}{\mu} \right)^{1-\varepsilon} [1 - A_k(\varepsilon)]^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} & \text{si } \varepsilon \neq 1 \\ \left[1 - \prod_{k=1}^K [1 - A_k(1)]^{q_k} \right] & \text{si } \varepsilon = 1 \end{cases} \\
A(\varepsilon)_B = \begin{cases} \left[1 - \left[\sum_{k=1}^K q_k \left(\frac{\mu_k}{\mu} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} \right] & \text{si } \varepsilon \neq 1 \\ \left[1 - \prod_{k=1}^K \left[\frac{\mu_k}{\mu} \right]^{q_k} \right] & \text{si } \varepsilon = 1 \end{cases}
\end{cases}$$

Les indicateurs d'inégalité ne sont pas tous décomposables. Parmi ceux qui le sont, certains sont additivement décomposables tandis que d'autres le sont multiplicativement. Certes la décomposabilité en sous-groupes qu'elle soit additive ou multiplicative est une propriété souhaitable aux indicateurs d'inégalité. Toutefois, il faut signaler qu'il n'existe pas un indicateur d'inégalité a priori meilleur que les autres. Il existe tout simplement plusieurs indicateurs qui mesurent différemment l'inégalité d'éducation. Ils sont donc complémentaires. De plus, chacun de ces indicateurs peut évaluer l'inégalité pour plusieurs types de variables éducatives : les ressources financières en éducation, notamment les dépenses par élève, les ressources matérielles (ratio élèves par classe), les ressources humaines (ratio élèves par enseignant), les résultats scolaires, notamment les compétences cognitives des apprenants (scores des élèves), l'accès à l'éducation (comme le taux d'analphabétisme, le niveau maximum atteint, le taux de scolarisation), le processus d'éducation (taux d'achèvement, de redoublement, d'abandon, de promotion). Toutefois, la construction de ces indicateurs à une échelle mondiale par rapport à l'une de ces variables éducatives reste fortement contrainte par la disponibilité de données harmonisées et comparables par rapport à la répartition de cette dernière.

Tableau 3: Quelques travaux sur la décomposition des inégalités éducatives selon différents critères de décomposition

Auteurs	Echantillon d'analyse	Les trois critères d'un indice d'inégalité d'éducation			Décomposition		Indices d'inégalité
		Variables d'éducation	Population d'intérêt	Entité d'observation	Critère de décomposition	Nombre de sous-groupes	
Zhang et Li (2002)	76 pays, 1960-1990	Moyenne d'années de scolarisation	Ensemble de pays	Pays	Niveau de développement	2	Indice de Gini
					Genre	2	
Sahn et Stifel (2003)	23 pays africains	Niveau d'éducation maximum atteint de la population âgée de 15-40 ans	Pays	Individus	Milieu	2	Indice de Theil
Mesa (2007)	Philippines, 2000	Moyenne d'années de scolarisation	Pays	Provinces	Niveau de pauvreté	2	Indice de Gini
	12 régions, Philippines, 2000	Niveau d'éducation maximum atteint	Région	Individus	Genre	2	
Sahn et Younger (2007)	49 pays en 2003	Score des élèves en maths & Sciences	Ensemble de pays	Individus	Pays	49	GE(0), GE(1), GE(2)
Corcoran et Evans (2007)	États-Unis, 1972-2004	Dépenses moyennes par élève	Pays (États-Unis)	Ecoles	États	50	Indice de Theil
Qian et Smyth (2008)	Chine, 1990-2000	Moyenne d'années de scolarisation	Pays	Provinces	Type de province côtières / intérieures	2	Indice de Gini
Lin et Yang (2009)	Taiwan, 1976-2004	Moyenne d'années de scolarisation	Pays	Individus	Tranches d'âges	11	Indice de Theil
Foldvari et van Leeuwen (2010)	28 pays de l'Europe	Moyenne d'années de scolarisation	Ensemble de pays de l'Europe	Individus	Pays	28	Indice de Gini
Gustave et Joubert (2012)	12 régions, Cameroun	Niveau d'éducation maximum atteint	Région	Individus	Niveau de pauvreté	2	Indice de Gini
					Genre	2	
Adrogue (2013)	Argentine, 2000	Indice composite de la qualité de l'école	Pays	Ecoles	Province	23	Indice de Theil
Agrawal (2014)	Inde, 1993-2009	Niveau d'éducation maximum atteint	Pays	Individus	Milieu	2	Indice de Gini
					États	18	
Yang et al. (2014)	Chine, 2004	Niveau d'éducation maximum atteint	Pays	Individus	Milieu	2	Indice de Gini
					Genre	2	
					Régions économiques	3	
					Niveau de revenu	5	
					Tranches d'âges	9	
Umar et al. (2014a)	Nigéria, 2010	Niveau d'éducation maximum atteint	Pays	Individus	Milieu	2	Indice de Theil
					Régions	6	
Borooah et Knox (2015)	Irlande, 2013	Score en anglais & mathématiques	Pays	Ecoles	Type des écoles	2	GE(0) (DLM)
Zamora et Dorado (2015)	Philippines, 1990-2010	Niveau d'éducation maximum atteint	Pays	Individus	Milieu	2	Indice de Gini
Tan et al. (2015)	État de Sabah, Malaisie 2009-2013	Moyenne des élèves (Fin du cycle secondaire)	État de Sabah (Malaisie)	Individus	Ecoles	188	Indice de Theil
					Milieu	2	

Source : Auteur.

1.2.4. Les bases mondiales des inégalités éducatives existantes

La construction d'une base mondiale sur les inégalités éducatives requiert l'existence de données sur la répartition de l'éducation par rapport à la variable considérée tout en ayant la même structure à la fois par rapport aux entités d'observation et la population d'intérêt, et ce, pour l'ensemble des pays considérés. Plusieurs auteurs conviennent qu'un tel exercice, de nature délicate, justifie l'absence de ces bases mondiales jusqu'il y a peu. La littérature sur les inégalités éducatives à l'échelle mondiale témoigne que le niveau de scolarité maximum atteint semble être la seule variable d'éducation qui a pu vérifier l'ensemble de ces critères pour un large panel de pays.

Ainsi, Johnes (2006) rapporte "The distribution of education is not a straightforward thing to measure. It requires information not only about the proportions of the current cohort of young people who leave education after completing various levels, but also similar information for every generation still alive. For many countries, this extent of detail in data which are quasi historic in nature has become available only very recently". De même, Meschi et Scervini (2014) notent que "the effects of the distribution of educational attainment have not been so widely studied... empirical evidence on this topic remains scarce, possibly owing to the lack of data comparing educational inequality across countries".

Le tableau n°4 présente un aperçu chronologique des bases mondiales sur les inégalités éducatives. Ainsi, Lopez et al. (1998) ont calculé un indice de Gini de l'éducation pour un échantillon de 12 pays¹⁸ entre 1970 et 1995. Par la suite, ils ont étendu leur version pour 20 pays. En utilisant ce même indice, Thomas et al. (2001) ont été les premiers à construire une base sur les inégalités d'éducation pour un large panel de (85) pays pour la période 1960-1990. Thomas et al. (2003) ont par la suite élargi leur base jusqu'à 140 pays de 1960 à 2000 et ont rajouté aussi l'indice de Theil de l'éducation. De même en utilisant l'indice de Gini de l'éducation, Castelló et Doménech (2002), Checchi (2004) et Lim et Tang (2008) ont construit une base de données comprenant respectivement 108 pays de 1960 à 2000 (935 observations), 117 pays de 1960 à 1995 (848 observations) et 99 pays de 1960 à 2000 (891 observations).

Les bases mondiales sur les inégalités éducatives sont toutes mesurées à partir de bases présentant la répartition de la population selon le niveau d'éducation maximum atteint. Toutefois, même s'il s'agit de la même variable, les sources mobilisées pour la construction de celle-ci diffèrent. Ainsi, Lopez et al. (1998, 1999), Thomas et al. (2001) et Checchi (2004) se sont appuyés sur la base de Barro et Lee (1996) tandis que Castelló et Doménech (2002), Thomas et al. (2003) et Lim et Tang (2008) se sont basés sur celle de Barro et Lee (2001). Araujo et al. (2004) et Meschi et Scervini (2014) se sont appuyés sur des bases d'enquêtes ménages, tandis que d'autres auteurs comme Castelló (2010a), Morrisson et Murtin (2013), Cuaresma et al. (2013), Földvári et Van Leeuwen (2014) et Sauer et Zagler (2014) ont utilisé des bases plus récentes comme celles de Cohen et Soto (2007), Lutz et al. (2007) ou bien celle de Barro et Lee (2013).

¹⁸ Les 12 pays de la base sont le Brésil, le Chili, la Chine, la Colombie, l'Inde, la Corée, le Mexique, la Malaisie, le Pérou, le Philippines, la Thaïlande et le Venezuela.

Tableau 4: Chronologie des bases mondiales sur les inégalités d'éducation

<i>Auteurs</i>	<i>Nombre de pays</i>	<i>Base de la répartition de l'éducation</i>	<i>Nombre de niveaux d'éducation</i>	<i>Période</i>	<i>Population</i>
Lopez et al. (1998)	12	BL (1996)	7	1970-1995	15+
Lopez et al. (1999)	20	BL (1996)	7	1970-1995	15+
Thomas et al. (2001)	85	BL (1996)	7	1960-1990	15+
Castelló et Doménech (2002)	108	BL (2001)	4	1960-2000	15+
Thomas et al. (2003)	140	BL (2001)	7	1960-2000	15+
Checchi (2004)	117	BL (1996)	4	1960-1995	25+
Araujo et al. (2004)	124	Enquêtes ménages	-	~ 2000	ATE
Lim et Tang (2008)	99	BL (2001)	7	1960-2000	25+
Castelló (2010a)	83	BL (2010)	4	1960-2000	25+
Morrisson et Murtin (2013)	78	MM (2009)	7	1870-2010	15-64 / 15+
Cuaresma et al. (2013)	175	Lutz et al. (2007)	4	1960-2010	Groupes d'âge
Földvári et Van Leeuwen (2014)	92 / 109	CS (2007) / BL (2001)	4	1960-2000	15+
Meschi et Scervini (2014)	31	Enquêtes ménages*	-	-	13 cohortes
Sauer et Zagler (2014)	134	BL (2013)	6	1955-2010	Groupes d'âge

ATE: Les adultes ayant terminé leurs études, 15+: Population âgée de 15 ans et plus, 25+: Population âgée de 25 ans et plus, ~2000 :

L'année la plus proche de 2000 pour laquelle les données d'enquête sur les ménages sont disponibles , BL: Barro et Lee, CS: Cohen et Soto, MM: Morisson et Murtin. * L'enquête sociale européenne, Statistiques de l'UE sur le revenu et les conditions de vie, l'Enquête internationale sur l'alphabétisation des adultes, Programme international d'enquêtes sociales.

Source : Auteur.

Tableau 5: Chronologie des bases de données sur le niveau d'éducation maximum atteint

<i>Auteurs</i>	<i>Nombre de pays</i>	<i>Période</i>
Psacharopoulos et Arriagada (1986)	99	1960-1980
Kyriacou (1991)	121	1965-1985
Lau et al. (1991)	58	1960-1986
Barro et Lee (1993)	129	1960-1985
Nehru et al. (1995)	85	1960-1987
Ahuja et Filmer (1996)	71	1985-1995
Barro et Lee (1996)	126	1960-1990
De la Fuente et Domenech (2000)	21 {OCDE}	1960-1990
Barro et Lee (2001)	142	1960-2000
De la Fuente et Domenech (2006)	21 {OCDE}	1960-1990
Cohen et Soto (2007)	95	1960-2000
Lutz et al. (2007)	120	1970-2000
Morrisson et Murtin (2009)	74	1870-1960
Barro et Lee (2013)	146	1950-2010
De la Fuente et Domenech (2015)	22 {OCDE}	1960-2010

Source : Auteur.

Plusieurs auteurs ont essayé de compiler des bases de données présentant la répartition de la population selon le niveau d'éducation maximum atteint (tableau n°5). L'une des premières tentatives a été faite par Psacharopoulos et Arriagada (1986) qui, à partir de plusieurs recensements nationaux, ont rassemblé des données sur la composition de la population âgée de plus de 25 ans selon six niveaux d'éducation. Depuis, de nombreux progrès ont été accomplis. Ainsi, la base de Barro et Lee (2013) présente un ensemble de données relatives à 146 pays par groupe d'âge et par genre de 1950 à 2010 sur des intervalles de cinq ans et qui donne à cet égard le pourcentage de la population – par tranche d'âge – ayant atteint, de façon achevée ou non, chacun des trois cycles d'enseignement (primaire, secondaire et supérieur).

Hormis cette erreur de pondération dans le calcul de la moyenne d'années de scolarisation, qui a été signalée dans le quatrième exemple de la première section – et qui peut être corrigée – ces auteurs ont sensiblement amélioré la précision de leurs estimations (les répartitions selon le niveau de scolarité). Barro et Lee ont utilisé exclusivement des données de recensements ventilées par groupe d'âge, ainsi que de nouvelles estimations des taux de mortalité et des taux d'achèvement différenciés selon l'âge et le niveau d'éducation par rapport à celles des versions antérieures (Barro et Lee 1993, 1996, et 2001).

Tableau 6: Comparaison entre les bases de données de Cohen et Soto (2007) et de Barro et Lee (2013)

<i>Critère de comparaison</i>	<i>Cohen and Soto (2007)</i>	<i>Barro and Lee (2013)</i>
Sources de données	Données de l'OCDE et l'UNESCO	Données de l'UNESCO
Nombre de pays	95	146
Nombre de pays développés (PD)	22	24
Nombre de recensements utilisés pour les PD	8	119
Nombre des sources de données utilisées (PD)	48	119
Nombre de pays en voie de développement (PVD)	73	122
Nombre de recensements utilisés pour les (PVD)	51	392
Nombre des sources de données utilisées (PVD)	70	392
Variation des taux de mortalité	Par groupes d'âge	Par groupes d'âge et niveau d'éducation
Variations dans les durées des cycles d'enseignement	homogène	Par groupes d'âge et dans le temps
Taux de fiabilité*	0.88	0.90

* Le taux de fiabilité a été utilisé par Krueger et Lindahl (2001) pour tester la qualité des données du capital humain dans les études transnationales.
Auteur.

Nous présentons dans le tableau n°6 une comparaison de la base de Barro et Lee (2013) avec celle de Cohen et Soto (2007). Cohen et Soto (2007) combinent deux sources différentes. Il s'agit des recensements nationaux rassemblés par l'UNESCO et des enquêtes des pays de l'OCDE. Barro et Lee (2001, 2013) ont montré qu'il existe une différence importante entre les données de l'OCDE et celles de l'UNESCO qui combinées, peuvent engendrer des incohérences considérables.

En effet, les données de l'OCDE proviennent essentiellement des enquêtes auprès des ménages qui sont bien évidemment moins robustes que les recensements. De plus, ces dernières ne sont disponibles qu'à partir des années 1990. Barro et Lee (2013) utilisent uniquement les recensements comme sources de données.

Comme le montre le tableau n°6, Cohen et Soto utilisent seulement 51 recensements sur les 70 sources de données utilisées pour 73 pays en développement et uniquement 8 recensements sur les 48 sources de données utilisées pour 22 pays développés. Tandis que Barro et Lee (2013) utilisent respectivement 392 et 199 recensements pour 122 pays en développement et 24 pays développés.

Il faut ajouter que le taux de fiabilité des estimations de Cohen et Soto (2007) est relativement moins élevé que celui de Barro et Lee (2013) (0,88 contre 0,90). De plus, malgré le fait que Cohen et Soto (2007) considèrent des taux de mortalité différenciés selon les groupes d'âge, ils ne tiennent pas compte de leurs variations selon les différents niveaux d'enseignement. Ils utilisent aussi des durées des cycles d'enseignement invariantes dans le temps contrairement à Barro et Lee (2013) qui en plus de la différenciation selon les groupes d'âge, prennent des taux de mortalité et des durées des cycles d'enseignement qui changent selon les différents niveaux d'enseignement et dans le temps.

1.3. Proposition d'une nouvelle base mondiale sur les inégalités éducatives

Nous proposons dans ce qui suit une nouvelle base mondiale sur les inégalités éducatives. Elle couvre 146 pays observés tout au long de 60 années (1950-2010), et propose une estimation de l'inégalité d'éducation aussi bien pour les tranches d'âge agrégées que pour les groupes d'âge de 5 ans. Par rapport à l'ensemble des bases existantes, nos calculs tendent à améliorer les estimations en particulier lorsqu'il s'agit de comparaisons internationales et ce, pour les raisons suivantes.

- Premièrement, comme la qualité de l'estimation de la répartition de l'éducation selon les différents niveaux d'éducation contribue à la précision de l'estimation de l'indice d'inégalité, nous avons utilisé la toute dernière base de Barro et Lee (2013) {version 2.0} qui, comme on l'a dit précédemment, offre des estimations plus précises et actualisées sur la répartition de l'éducation selon le niveau maximum atteint et présente des avantages spécifiques par rapport aux autres bases mondiales¹⁹.

- Deuxièmement, nous avons eu recours à sept niveaux d'éducation. Les bases de Castelló et Doménech (2002), Checchi (2004), Castelló (2010a), Cuaresma et al. (2013) et Van Leeuwen et Földvári (2014) ont toutes été construites en s'appuyant uniquement sur quatre niveaux d'éducation. Ces auteurs ne font pas de distinction entre ceux qui ont effectivement achevé un cycle d'enseignement et ceux qui ne l'ont pas achevé puisqu'ils utilisent des données agrégées par cycle. Dans ce

¹⁹ Il faut signaler que l'erreur de pondération qu'on a pu trouver dans la base de Barro et Lee (2013) et qui sera explicitée après ne remet pas en question la qualité des données sur la répartition de l'éducation mais concerne plutôt les données sur les moyennes d'années de scolarisation de la population âgée de 15 ans et plus et de 25 ans et plus. Pour calculer l'indice d'inégalité nous avons utilisé de préférence les données sur la répartition de l'éducation et nous avons mené les corrections nécessaires pour les moyennes d'années de scolarisation.

cas, l'inégalité estimée concerne uniquement l'inégalité inter-cycles alors que l'inégalité intra-cycles est de ce fait ignorée.

- Troisièmement, comme nous exploitons des données ventilées par tranches d'âge, nous prenons en compte la variation dans les durées des cycles d'enseignement. Il faut souligner que l'ensemble des bases mondiales élaborées jusqu'à présent pour quantifier les inégalités éducatives ont toutes utilisé des durées des cycles d'enseignement fixes dans le temps. Tenir compte de ces changements nécessite au préalable de disposer de données désagrégées par tranches d'âge. La population âgée de 15 ans et plus est constituée d'individus, appartenant à différentes classes d'âge, n'ayant pas forcément passé au cours du temps par le même système d'éducation. En effet, durant la période 1950-2010, des restructurations du système éducatif ont été entreprises dans la plupart des pays impliquant des changements dans la durée totale de la scolarité aux cycles primaire collégial et secondaire. Nous avons recensé 99 pays sur les 146 de la base de Barro et Lee (2013) qui ont connu des changements dans ces durées, dont 31 pays ont eu plus de trois changements. Bien entendu, ne pas tenir compte de ces changements risque de modifier significativement les estimations des indices d'inégalité. Ceci impacte aussi bien l'étendue que la répartition du nombre d'années d'études effectuées par les individus de ces différentes tranches d'âge.

- Finalement, la base de données que nous proposons est la première à offrir un indice de Gini d'éducation ventilé en tranches d'âge de 5 ans, par genre, pour 146 pays, pour des groupes agrégés de pays et/ou d'intervalles d'âge, tout en tenant compte de la variation et de l'hétérogénéité au sein des groupes agrégés.

Nous présentons dans ce qui suit la méthodologie adoptée pour construire l'indice de Gini de l'éducation basé sur la répartition de la population selon le niveau maximum atteint dans notre base mondiale sur les inégalités éducatives.

1.3.1. Méthodologie

- ***L'indice de Gini de l'éducation d'un groupe d'âge quinquennal***

L'indice de Gini demeure le coefficient le plus utilisé pour mesurer les inégalités dans la répartition des revenus. Toutefois, comme on l'a souligné précédemment, il peut être adapté à toutes les variables de l'éducation. Les indices de Gini – de l'éducation et du revenu – sont tous les deux égaux à la moitié de la moyenne des écarts absolus entre les observations prises deux à deux rapporté sur la moyenne des observations. De même ces deux indices peuvent aussi être définis en faisant référence à la courbe de Lorenz. La seule différence entre les deux est que le niveau de scolarité maximum atteint est une variable discrète, tandis que le revenu est le plus souvent considéré comme une variable continue mais peut toutefois être adapté au cas discret en utilisant des regroupements par intervalle de revenu. A cet égard, Lambert (2001) a présenté la formule de l'indice de Gini du revenu pour le cas discret²⁰. Afin d'harmoniser les formules de ces deux indices, on peut, à partir de la formule proposée par Lambert (2001), retrouver celle suggérée par Thomas et al. (2001) comme suit :

²⁰ Voir Giorgi et Gigliarano (2017) pour plus de détails sur les formulations les plus populaires de l'indice de Gini dans le cas discret.

$$\begin{aligned}
Gini &= \frac{\sum_{\alpha=1}^N \sum_{\beta=1}^N |y_{\alpha} - y_{\beta}|}{2N^2 \mu} \\
&= \frac{\sum_{i=1}^n \sum_{j=1}^n \left(\frac{1}{N^2}\right) \sum_{h=1}^{n_i} \sum_{k=1}^{n_j} |y_{ih} - y_{jk}|}{2\mu} \\
&= \frac{\sum_{i=1}^n \sum_{j=1}^n \left(\frac{n_i}{N}\right) \left(\frac{n_j}{N}\right) |y_i - y_j|}{2\mu} \\
&= \frac{\sum_{i=1}^n \sum_{j=1}^n p_i |y_i - y_j| p_j}{2\mu} \\
&= \frac{1}{\mu} \sum_{i=1}^n \sum_{j=1}^{i-1} p_i |y_i - y_j| p_j
\end{aligned}$$

Afin d'homogénéiser l'ensemble des formules qui vont suivre, on peut réécrire la formule de l'indice de Gini de l'éducation comme suit :

$$Gini = \frac{\sum_{i=1}^n \sum_{j=1}^n p_i |y_i - y_j| p_j}{2\mu} = \frac{\sum_{i=1}^n \sum_{j=1}^n \left(\frac{n_i}{N}\right) |y_i - y_j| \left(\frac{n_j}{N}\right)}{2 \sum_{i=1}^n \left(\frac{n_i}{N}\right) y_i}$$

D'où la formule suivante :

$$Gini = \frac{\sum_{i=1}^n \sum_{j=1}^n n_i |y_i - y_j| n_j}{2N \sum_{i=1}^n n_i y_i}$$

L'indice de Gini de l'éducation d'un groupe d'âge quinquennal est donc :

$$Gini_{c,t}^a = \frac{\sum_{i=1}^n \sum_{j=1}^n n_{c,i,t}^a |y_{c,i,t}^a - y_{c,j,t}^a| n_{c,j,t}^a}{2 N_{c,t}^a \sum_{i=1}^n n_{c,i,t}^a y_{c,i,t}^a}$$

où :

- $Gini_{c,t}^a$ est l'indice de Gini de l'éducation du groupe d'âge "a" appartenant au pays "c" à la date "t".
- a=1 pour le groupe d'âge 15–19 ans, a=2 pour le groupe d'âge 20–24 ans, ..., et a=13 pour le groupe d'âge 75 et plus.
- n est le nombre des niveaux d'éducation retenus qui correspond dans notre cas à 7
- i et j sont les niveaux d'éducation tel que: i=1 pour les illettrés, i=2 pour le primaire inachevé, i=3 pour le primaire achevé, i=4 pour le secondaire inachevé, i=5 pour le secondaire achevé, i=6 pour le supérieur inachevé, and i=7 pour le supérieur achevé.

- $n_{c,i,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" ayant atteint le niveau d'éducation "i" à la date "t".
- $N_{c,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" à la date "t" ($N_{c,t}^a = \sum_{i=1}^n n_{c,i,t}^a$).
- $y_{c,i,t}^a$ est le nombre d'années de scolarisation cumulées de la tranche d'âge "a" du pays "c" pour atteindre le niveau d'éducation "j" à la date "t".

Il faut signaler que les $y_{c,j,t}^a$ ne sont pas explicitement disponibles dans la base de Barro and Lee (2013). Nous les avons calculés en utilisant notre propre méthodologie (voir section 1.3.2).

- ***L'indice de Gini de l'éducation pour des groupes agrégés et l'erreur de Barro et Lee***

Afin de calculer l'indice de Gini de l'éducation pour des grands groupes d'âge (15 ans et plus, 25 ans et plus, [15, 65] ans, ...), la plupart des auteurs, même ceux qui ont utilisé la base de Barro et Lee (2013) comme Sauer et Zagler (2014), ne prennent pas en considération la variation par tranches d'âge des durées des cycles d'enseignement. Par exemple pour calculer l'indice de Gini de l'éducation pour la population âgée de 15 ans et plus, ils supposent des « $y_{c,j,t}^a$ » invariants à la fois par tranche d'âge et dans le temps ($y_{c,j,t}^{15+} = y_{c,j}^a = y_{c,j} \forall a \in [1 - 13], \forall t \in [1950 - 2010]$). Ainsi la formule qu'ils utilisent est la suivante :

$$\left\{ \begin{array}{l} Gini_{c,t}^{15+} = \frac{\sum_{i=2}^n \sum_{j=1}^{i-1} p_{c,i,t}^{15+} |y_{c,i,t}^{15+} - y_{c,j,t}^{15+}| p_{c,j,t}^{15+}}{\sum_{j=1}^n p_{c,j,t}^{15+} y_{c,j,t}^{15+}} \text{ où } p_{c,j,t}^{15+} = \sum_{a=1}^A l_{c,t}^a p_{c,j,t}^a \text{ Avec } l_{c,t}^a = \frac{Pop_{c,t}^a}{Pop_{c,t}^{15+}} \\ \text{sous l'hypothèse que } y_{c,j,t}^{15+} = y_{c,j}^a = y_{c,j} \forall a \in [1 - 13] \forall t \in [1950 - 2010] \end{array} \right.$$

Toutefois, même s'il s'agit du même pays "c" à la même date "t", les durées des cycles d'enseignement peuvent être différentes d'une tranche d'âge à l'autre. De manière générale, sur les 146 pays de notre échantillon, 99 pays ont recouru à des changements dans les durées des cycles d'enseignement.

Une première approche serait alors de calculer une moyenne des « $y_{c,j,t}^a$ » de manière à tenir compte de ces éventuelles variations par rapport aux différents groupes d'âge. Ainsi, comme $p_{c,j,t}^{15+} = \sum_{a=1}^A l_{c,t}^a p_{c,j,t}^a$, on pourrait a priori penser que $y_{c,j,t}^{15+}$ serait de même égale à $\sum_{a=1}^A l_{c,t}^a y_{c,j,t}^a$.

Comme on l'a déjà signalé précédemment, Barro et Lee (2013) n'ont pas utilisé la pondération adéquate dans le calcul des durées $y_{c,j,t}^{15+}$. En effet, après avoir déduit les durées $y_{c,j,t}^a$ à partir de leur base de données²¹, nous avons constaté que nos moyennes d'années de scolarisation par tranches d'âge quinquennales ($MAS_{c,t}^a$) correspondent parfaitement à celles calculées par Barro et Lee (2013). Toutefois, lorsqu'il s'agit de calculer la moyenne d'années de scolarisation de la tranche d'âge agrégée de 15 ans et plus, en appliquant la formule utilisée dans leur article $MAS_{c,t}^{15+} = \sum_{a=1}^A l_{c,t}^a MAS_{c,t}^a$ (la moyenne des moyennes d'années de scolarisation par

²¹ La démarche de calcul des $y_{c,j,t}^a$ sera présentée dans la section 1.3.2.

tranches d'âge quinquennales) ainsi que leurs propres données (répartition de la population selon les niveaux d'éducation) on ne retrouve pas les mêmes résultats.

Il s'avère que, pour la population âgée de 15 ans et plus, les auteurs ont plutôt utilisé une autre formule tout aussi valable qui est égale à la somme des produits entre les durées et les proportions $MAS_{c,t}^{15+} = \sum_{j=1}^n p_{c,j,t}^{15+} y_{c,j,t}^{15+}$. Auquel cas, le problème proviendrait des durées agrégées $y_{c,j,t}^{15+}$. A priori, les auteurs se sont appuyés sur une pondération qui correspondrait à celle utilisée pour les $p_{c,j,t}^{15+}$ ($p_{c,j,t}^{15+} = \sum_{a=1}^A l_{c,t}^a p_{c,j,t}^a$). Afin de vérifier cela, nous avons recalculé les durées agrégées en utilisant cette dernière pondération tel que $\tilde{y}_{c,j,t}^{15+} = \sum_{a=1}^A l_{c,t}^a y_{c,j,t}^a$. Nous retrouvons alors les mêmes valeurs des moyennes d'années de scolarisation des 15 ans et plus que celles des auteurs²².

En partant des formules de la moyenne d'années de scolarisation de la tranche d'âge "a" ($MAS_{c,t}^a = \sum_{j=1}^n p_{c,j,t}^a y_{c,j,t}^a$) et de celle de la population âgée de 15 ans et plus ($MAS_{c,t}^{15+} = \sum_{a=1}^A l_{c,t}^a MAS_{c,t}^a = \sum_{j=1}^n p_{c,j,t}^{15+} y_{c,j,t}^{15+}$), on peut démontrer que l'agrégation des « $y_{c,j,t}^a$ » pour l'ensemble de la population âgée de 15 ans et plus s'écrit comme suit :

$$y_{c,j,t}^{15+} = \sum_{a=1}^A \left[\frac{l_{c,t}^a p_{c,j,t}^a}{\sum_{b=1}^A l_{c,t}^b p_{c,j,t}^b} \right] y_{c,j,t}^a$$

Certes, l'erreur de pondération faite par Barro et Lee (2013) peut, pour certains pays, ne pas avoir d'impact si les $y_{c,j,t}^a$ restent les mêmes pour l'ensemble des tranches d'âge pour une même année. Toutefois comme les systèmes éducatifs de la majorité des pays ont subi des aménagements quant aux durées des cycles d'enseignement durant 60 ans, cette erreur peut changer significativement les résultats et ceci peut être illustré à travers l'exemple qui suit.

Les tableaux n°7 et n°8 présentent le cas de la Bulgarie en 1990 {c=120,t=1990}. Le tableau n°7 fournit la répartition de la population de la Bulgarie en 1990 par tranche d'âge selon le niveau d'éducation maximum atteint « $p_{c,j,t}^a$ » ainsi que les poids de chaque tranche d'âge dans la population âgée de 15 ans et plus « $l_{c,t}^a$ ». Le tableau n°8 présente les nombres d'années cumulées de scolarisation par tranche d'âge et selon chaque niveau d'éducation « $y_{c,j,t}^a$ » ainsi que la moyenne d'années de scolarisation par tranche d'âge. La différence entre les deux pondérations est ainsi présentée dans le tableau n°8.

Par exemple si on prend, pour $j=3$, $y_{c,3,t}^{15+}$: selon Barro et Lee (2013), sa valeur est égale à $\sum_{a=1}^A l_{c,t}^a y_{c,3,t}^a = 5.4$ alors qu'elle doit plutôt être égale à $\sum_{a=1}^A \left[\frac{l_{c,t}^a p_{c,3,t}^a}{\sum_{b=1}^A l_{c,t}^b p_{c,3,t}^b} \right] y_{c,3,t}^a = 4.8$. De même pour les autres $y_{c,j,t}^{15+}$, ce qui implique que la moyenne d'années de scolarisation de la population âgée de 15 ans et plus peut passer de 7.96 à 8.42 (différence de 0.46). Cette différence est relativement importante, puisqu'une moyenne d'années de scolarisation n'augmente en moyenne que d'une année de scolarisation tous les dix ans. D'où l'ampleur du biais dû à l'utilisation d'une pondération non adéquate.

²² On a refait les calculs pour les versions précédentes de la base {1.0, 1.1, 1.2, 1.3 et 2.0} et on obtient les mêmes résultats.

Tableau 7: Répartition de la population par tranche d'âge selon le niveau d'éducation maximum atteint pour le cas de la Bulgarie en 1990

Tranche d'âge	Population	Poids de la tranche d'âge (l^a)	Proportion de la population selon le niveau d'éducation maximum atteint (en %) ($p_{c,j,t}^a$) {c=120,t=1990}						
			Non scolarisés (p_1^a)	Primaire inachevé (p_2^a)	Primaire achevé (p_3^a)	Secondaire inachevé (p_4^a)	Secondaire achevé (p_5^a)	Supérieur inachevé (p_6^a)	Supérieur achevé (p_7^a)
15-19	641	9.24	1.90	6.40	63.50	22.76	5.24	0.09	0.01
20-24	585	8.43	1.40	2.80	17.30	14.97	57.13	4.26	2.04
25-29	580	8.36	1.40	2.70	17.40	12.35	46.35	6.33	13.37
30-34	606	8.74	1.40	2.70	17.40	13.59	45.11	6.33	13.37
35-39	625	9.01	1.40	3.40	25.70	14.34	34.06	6.72	14.18
40-44	638	9.20	1.40	3.40	25.70	16.46	31.94	6.72	14.18
45-49	527	7.60	2.70	6.80	37.90	16.70	19.00	5.37	11.33
50-54	513	7.40	2.70	6.80	37.90	18.73	16.97	5.37	11.33
55-59	560	8.07	5.60	16.30	42.80	14.26	9.64	3.63	7.67
60-64	529	7.63	5.60	16.30	42.80	14.86	9.04	2.75	8.55
65-69	474	6.83	12.00	32.30	36.70	8.55	3.75	1.61	4.99
70-74	240	3.46	12.00	32.30	36.70	8.95	3.35	1.09	5.51
75 ans et plus	418	6.03	12.00	32.30	36.70	9.45	2.85	1.25	5.35
15 ans et plus	6936	-	4.03	9.13	35.00	18.44	20.19	4.96	8.11

Source : Base de données de Barro et Lee (2013)

Tableau 8: Moyenne d'année de scolarisation et nombre d'années cumulées de scolarisation par tranche d'âge pour le cas de la Bulgarie en 1990

Tranche d'âge	Moyenne d'années de scolarisation	Nombre d'années cumulées de scolarisation ($y_{c,j,t}^a$) {c=120,t=1990}						
		Non scolarisés (y_1^a)	Primaire inachevé (y_2^a)	Primaire achevé (y_3^a)	Secondaire inachevé (y_4^a)	Secondaire achevé (y_5^a)	Supérieur inachevé (y_6^a)	Supérieur achevé (y_7^a)
15-19	5.08	0.0	2.0	4.0	8.0	11.0	13.0	15.0
20-24	10.06	0.0	4.0	8.0	9.5	11.0	13.0	15.0
25-29	10.60	0.0	4.0	8.0	9.5	11.0	13.0	15.0
30-34	10.58	0.0	4.0	8.0	9.5	11.0	13.0	15.0
35-39	10.30	0.0	4.0	8.0	9.5	11.0	13.0	15.0
40-44	8.84	0.0	2.0	4.0	7.5	11.0	13.0	15.0
45-49	8.88	0.0	2.0	4.0	10.0	14.0	16.0	18.0
50-54	8.80	0.0	2.0	4.0	10.0	14.0	16.0	18.0
55-59	6.78	0.0	2.0	4.0	10.0	14.0	16.0	18.0
60-64	6.77	0.0	2.0	4.0	10.0	14.0	16.0	18.0
65-69	4.65	0.0	2.0	4.0	10.0	14.0	16.0	18.0
70-74	4.64	0.0	2.0	4.0	10.0	14.0	16.0	18.0
75 ans et plus	4.62	0.0	2.0	4.0	10.0	14.0	16.0	18.0
15 ans et plus (Auteur)	7.96	0.0	2.2	4.8	9.3	11.6	14.1	16.3
15 ans et plus (BL2013)	8.42	0.0	2.7	5.4	9.4	12.4	14.4	16.4
Surestimation	0.46	0.0	0.5	0.6	0.1	0.8	0.3	0.1

Source : Calcul de l'auteur à partir de base de données de Barro et Lee (2013)

Ainsi, opter pour une solution qui consiste à calculer la moyenne des « $y_{c,j,t}^a$ » avec les bons poids pour obtenir « $y_{c,j,t}^{15+}$ » puis l'utiliser par la suite pour l'évaluation de l'indice de Gini de l'éducation de la population âgée de 15 ans et plus, semble être une approche pertinente. Toutefois, elle consiste à agréger les données avant de calculer l'indice de Gini du groupe d'âge agrégé et ne permet cependant pas de prendre en considération les inégalités inter-tranches d'âge ni d'exploiter la richesse des données désagrégées. Afin de tenir compte de cela, nous utilisons dans la construction de notre base mondiale des inégalités éducatives une autre approche qui consiste à exploiter toute la richesse de l'ensemble des données existantes en utilisant une formule plus adaptée :

$$Gini_{c,t}^G = \frac{\sum_{a \in G} \sum_{i \in G} \sum_{j=1}^n \sum_{i=1}^n n_{c,i,t}^a |y_{c,i,t}^a - y_{c',j,t}^{a'}| n_{c',j,t}^{a'}}{2 N_{c,t}^G \left[\sum_{a \in G} \sum_{i=1}^n n_{c,i,t}^a y_{c,i,t}^a \right]}$$

avec:

- $Gini_{c,t}^G$ l'indice de Gini de l'éducation du grand groupe d'âge "G" appartenant au pays "c" à la date "t".²³
- $a=1$ ($a'=1$) pour le groupe d'âge 15–19 ans, $a=2$ ($a'=2$) pour le groupe d'âge 20–24 ans, ..., et $a=13$ ($a'=13$) pour le groupe d'âge 75 et plus.
- i et j sont les niveaux d'éducation : $i=1$ pour les illettrés, $i=2$ pour le primaire inachevé, $i=3$ pour le primaire achevé, $i=4$ pour le secondaire inachevé, $i=5$ pour le secondaire achevé, $i=6$ pour le supérieur inachevé, and $i=7$ pour le supérieur achevé.
- $n_{c,i,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" ayant atteint le niveau d'éducation "i" à la date "t".
- $y_{c,i,t}^a$ traduit le nombre d'années de scolarisation cumulées de la tranche d'âge "a" du pays "c" pour atteindre le niveau d'éducation "j" à la date "t".

De même pour des groupes de pays, qu'ils soient considérés selon le niveau de développement (pays développés, pays en développement) ou selon la zone géographique (Moyen-Orient et Afrique du Nord, Afrique subsaharienne, Asie du Sud...), des auteurs comme Földvári et Van Leeuwen (2014) calculent une moyenne des indices de Gini par pays au lieu de calculer un indice de Gini sur la distribution de l'ensemble des pays appartenant au groupe en question et ignorent de ce fait, les inégalités inter-pays. Pour rendre compte du niveau d'inégalités pour un groupe de pays, nous calculons un indice de Gini sur la distribution de l'ensemble des pays :

$$Gini_{R,t}^a = \frac{\sum_{c \in R} \sum_{c' \in R} \sum_{i=1}^n \sum_{j=1}^n n_{c,i,t}^a |y_{c,i,t}^a - y_{c',j,t}^a| n_{c',j,t}^a}{2 N_{R,t}^a \left[\sum_{c \in R} \sum_{i=1}^n n_{c,i,t}^a y_{c,i,t}^a \right]} \quad (6)$$

avec :

²³ Par exemple G peut être la population âgée entre [15, 65] ans, [25-65] ans, [15-24], ou encore la population âgée de 15 ans et plus ou 25 ans et plus...

- $Gini_{R,t}^a$ l'indice de Gini de l'éducation du groupe d'âge "a" appartenant à la région "R" à la date "t".
- c (c') représente les pays appartenant à la région "R".
- a=1 (a'=1) pour le groupe d'âge 15–19 ans, a=2 (a'=2) pour le groupe d'âge 20–24 ans, ..., et a=13 (a'=13) pour le groupe d'âge 75 et plus.
- n est le nombre des niveaux d'éducation retenus qui correspond dans notre cas à 7
- i et j sont les niveaux d'éducation tel que: i=1 pour les illettrés, i=2 pour le primaire inachevé, i=3 pour le primaire achevé, i=4 pour le secondaire inachevé, i=5 pour le secondaire achevé, i=6 pour le supérieur inachevé, and i=7 pour le supérieur achevé.
- $n_{c,i,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" ayant atteint le niveau d'éducation "i" à la date "t".
- $N_{R,t}^a$ représente la taille de la population du groupe d'âge "a" appartenant à la région "R" à la date "t". ($N_{R,t}^a = \sum_{c \in R} \sum_{i=1}^n n_{c,i,t}^a = \sum_{c \in R} N_{c,t}^a$).
- $y_{c,i,t}^a$ indique le nombre d'années de scolarisation cumulées de la tranche d'âge "a" du pays "c" pour atteindre le niveau d'éducation "j" à la date "t".

On peut généraliser la formule du calcul de l'indice de Gini pour des groupes agrégés à la fois par pays et par grands groupes d'âge (par exemple, pour la population âgée de 15 ans et plus de la région MENA) tout en utilisant une formule qui, en plus de prendre en considération les inégalités inter-niveaux d'éducation des individus, tienne compte à la fois des inégalités inter-pays et inter-tranches d'âge :

$$Gini_{R,t}^G = \frac{\sum_{c \in R} \sum_{c' \in R} \sum_{a \in G} \sum_{a' \in G} \sum_{i=1}^n \sum_{j=1}^n n_{c,i,t}^a |y_{c,i,t}^a - y_{c',j,t}^{a'}| n_{c',j,t}^{a'}}{2 N_{R,t}^G \left[\sum_{c \in R} \sum_{a \in G} \sum_{i=1}^n n_{c,i,t}^a y_{c,i,t}^a \right]} \quad (7)$$

avec :

- $Gini_{R,t}^G$ l'indice de Gini de l'éducation du grand groupe d'âge "G" appartenant à la région "R" à la date "t".
- a=1 pour le groupe d'âge 15–19 ans, a=2 pour le groupe d'âge 20–24 ans, ..., et a=13 pour le groupe d'âge 75 et plus.
- c (c') représente les pays appartenant à la région "R".
- n est le nombre des niveaux d'éducation retenus qui correspond dans notre cas à 7
- i et j sont les niveaux d'éducation tel que: i=1 pour les illettrés, i=2 pour le primaire inachevé, i=3 pour le primaire achevé, i=4 pour le secondaire inachevé, i=5 pour le secondaire achevé, i=6 pour le supérieur inachevé, and i=7 pour le supérieur achevé.
- $n_{c,i,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" ayant atteint le niveau d'éducation "i" à la date "t".
- $N_{R,t}^G$ représente la taille de la population du grand groupe d'âge "G" appartenant à la région "R" à la date "t". ($N_{R,t}^G = \sum_{c \in R} \sum_{a \in G} \sum_{i=1}^n n_{c,i,t}^a = \sum_{c \in R} N_{c,t}^G = \sum_{c \in R} N_{R,t}^a$).

- $y_{c,i,t}^a$ indique le nombre d'années de scolarisation cumulées de la tranche d'âge "a" du pays "c" pour atteindre le niveau d'éducation "j" à la date "t".

- ***Les indices d'entropie généralisée et d'Atkinson de l'éducation***

En suivant la même approche adoptée pour le calcul de l'indice de Gini de l'éducation d'un groupe d'âge quinquennal d'un pays donné ou d'un groupe agrégé de pays et/ou de groupes d'âge, nous pouvons aussi calculer d'autres indices d'inégalité comme ceux appartenant à la famille d'entropie généralisée ou d'Atkinson²⁴ :

Les indices d'entropie généralisée et d'Atkinson de l'éducation sont :

$$GE(r)_{c,t}^a = \frac{N_{c,t}^a (r-1)}{r(r-1)} \left[\frac{\sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^r}{\left(\sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a \right)^r} - 1 \right] \text{ et } A(\varepsilon)_{c,t}^a = 1 - \frac{1}{\left[N_{c,t}^a \right]^{\left(\frac{\varepsilon}{1-\varepsilon} \right)}} \left[\frac{\left[\sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^{(1-\varepsilon)} \right]^{\left(\frac{1}{1-\varepsilon} \right)}}{\sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a} \right]$$

avec :

- $GE(r)_{c,t}^a$ l'indice d'entropie généralisée de l'éducation d'ordre r du groupe d'âge "a" appartenant au pays "c" à la date "t".
 - $A(\varepsilon)_{c,t}^a$ l'indice Atkinson de l'éducation d'ordre ε du groupe d'âge "a" appartenant au pays "c" à la date "t".
 - $a=1$ pour le groupe d'âge 15–19 ans, $a=2$ pour le groupe d'âge 20–24 ans, ..., et $a=13$ pour le groupe d'âge 75 et plus.
 - n est le nombre des niveaux d'éducation retenus qui correspond dans notre cas à 7
 - i et j sont les niveaux d'éducation tel que: $i=1$ pour les illettrés, $i=2$ pour le primaire inachevé, $i=3$ pour le primaire achevé, $i=4$ pour le secondaire inachevé, $i=5$ pour le secondaire achevé, $i=6$ pour le supérieur inachevé, and $i=7$ pour le supérieur achevé.
 - $n_{c,i,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" ayant atteint le niveau d'éducation "i" à la date "t".
 - $N_{c,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" à la date "t" ($N_{c,t}^a = \sum_{i=1}^n n_{c,i,t}^a$).
 - $y_{c,i,t}^a$ indique le nombre d'années de scolarisation cumulées de la tranche d'âge "a" du pays "c" pour atteindre le niveau d'éducation "j" à la date "t".
- ***Les indices d'entropie généralisée et d'Atkinson de l'éducation pour des groupes agrégés***

❖ **Grand Groupe d'âge d'un pays**

²⁴ L'annexe n°1 présente une synthèse de l'ensemble des formules qui ont été proposées pour le calcul des indices d'inégalité éducative.

$$\left\{ \begin{array}{l} GE(r)_{c,t}^G = \frac{N_{c,t}^{G(r-1)}}{r(r-1)} \left[\frac{\sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^r}{\left(\sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a \right)^r} - 1 \right] \\ A(\varepsilon)_{c,t}^G = 1 - \frac{1}{\left[N_{c,t}^G \right]^{\left(\frac{\varepsilon}{1-\varepsilon} \right)}} \frac{\left[\sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^{(1-\varepsilon)} \right]^{\left(\frac{1}{1-\varepsilon} \right)}}{\sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a} \end{array} \right.$$

avec :

- $GE(r)_{c,t}^G$ l'indice d'entropie généralisée de l'éducation d'ordre r du grand groupe d'âge "G" appartenant au pays "c" à la date "t".²⁵
- $A(\varepsilon)_{c,t}^G$ l'indice Atkinson de l'éducation d'ordre ε du grand groupe d'âge "G" appartenant au pays "c" à la date "t".
- $a=1$ ($a'=1$) pour le groupe d'âge 15–19 ans, $a=2$ ($a'=2$) pour le groupe d'âge 20–24 ans, ..., et $a=13$ ($a'=13$) pour le groupe d'âge 75 et plus.
- n le nombre des niveaux d'éducation retenus qui correspond dans notre cas à 7.
- $N_{c,t}^G$ représente la taille de la population du grand groupe d'âge "G" du pays "c" à la date "t" ($N_{c,t}^G = \sum_{a \in G} \sum_{i=1}^n n_{c,i,t}^a = \sum_{a \in G} N_{c,t}^a$)
- i et j sont les niveaux d'éducation tel que: $i=1$ pour les illettrés, $i=2$ pour le primaire inachevé, $i=3$ pour le primaire achevé, $i=4$ pour le secondaire inachevé, $i=5$ pour le secondaire achevé, $i=6$ pour le supérieur inachevé, and $i=7$ pour le supérieur achevé.
- $n_{c,i,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" ayant atteint le niveau d'éducation "i" à la date "t".
- $y_{c,i,t}^a$: le nombre d'années de scolarisation cumulées de la tranche d'âge "a" du pays "c" pour atteindre le niveau "j" à la date "t".

❖ **Groupe d'âge quinquennal d'une zone géographique**

$$\left\{ \begin{array}{l} GE(r)_{R,t}^a = \frac{N_{R,t}^{a(r-1)}}{r(r-1)} \left[\frac{\sum_{c \in R} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^r}{\left(\sum_{c \in R} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a \right)^r} - 1 \right] \\ A(\varepsilon)_{R,t}^a = 1 - \frac{1}{\left[N_{R,t}^a \right]^{\left(\frac{\varepsilon}{1-\varepsilon} \right)}} \frac{\left[\sum_{c \in R} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^{(1-\varepsilon)} \right]^{\left(\frac{1}{1-\varepsilon} \right)}}{\sum_{c \in R} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a} \end{array} \right.$$

²⁵ Par exemple G peut être la population âgée entre [15, 65] ans, [25-65] ans, [15-24], ou encore la population âgée de 15 ans et plus ou 25 ans et plus...

avec :

- $GE(r)_{R,t}^a$ l'indice d'entropie généralisée de l'éducation d'ordre r du groupe d'âge "a" appartenant à la région "R" à la date "t".
- $A(\varepsilon)_{R,t}^a$ l'indice Atkinson de l'éducation d'ordre ε du groupe d'âge "a" appartenant à la région "R" à la date "t".
- c (c') les pays appartenant à la région "R".
- $a=1$ ($a'=1$) pour le groupe d'âge 15–19 ans, $a=2$ ($a'=2$) pour le groupe d'âge 20–24 ans, ..., et $a=13$ ($a'=13$) pour le groupe d'âge 75 et plus.
- n est le nombre des niveaux d'éducation retenus qui correspond dans notre cas à 7
- i et j sont les niveaux d'éducation tel que: $i=1$ pour les illettrés, $i=2$ pour le primaire inachevé, $i=3$ pour le primaire achevé, $i=4$ pour le secondaire inachevé, $i=5$ pour le secondaire achevé, $i=6$ pour le supérieur inachevé, and $i=7$ pour le supérieur achevé.
- $n_{c,i,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" ayant atteint le niveau d'éducation "i" à la date "t".
- $N_{R,t}^a$ représente la taille de la population du groupe d'âge "a" appartenant à la région "R" à la date "t". ($N_{R,t}^a = \sum_{c \in R} \sum_{i=1}^n n_{c,i,t}^a = \sum_{c \in R} N_{c,t}^a$).
- $y_{c,i,t}^a$ indique le nombre d'années de scolarisation cumulées de la tranche d'âge "a" du pays "c" pour atteindre le niveau d'éducation "j" à la date "t".

❖ Grand Groupe d'âge d'une zone géographique

$$\left\{ \begin{array}{l} GE(r)_{c,t}^G = \frac{N_{R,t}^{G(r-1)}}{r(r-1)} \left[\frac{\sum_{c \in R} \sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^r}{\left(\sum_{c \in R} \sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a \right)^r} - 1 \right] \\ A(\varepsilon)_{R,t}^G = 1 - \frac{1}{\left[N_{R,t}^G \right]^{\left(\frac{\varepsilon}{1-\varepsilon} \right)}} \left[\frac{\sum_{c \in R} \sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^{(1-\varepsilon)} \right]^{\left(\frac{1}{1-\varepsilon} \right)}}{\sum_{c \in R} \sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a} \right] \end{array} \right.$$

avec :

- $GE(r)_{R,t}^G$ l'indice d'entropie généralisée de l'éducation d'ordre r du grand groupe d'âge "G" appartenant à la région "R" à la date "t".
- $A(\varepsilon)_{R,t}^G$ l'indice Atkinson de l'éducation d'ordre ε du grand groupe d'âge "G" appartenant à la région "R" à la date "t".
- $a=1$ pour le groupe d'âge 15–19 ans, $a=2$ pour le groupe d'âge 20–24 ans, ..., et $a=13$ pour le groupe d'âge 75 et plus.
- c (c') les pays appartenant à la région "R".
- n est le nombre des niveaux d'éducation retenus qui correspond dans notre cas à 7
- i et j sont les niveaux d'éducation tel que: $i=1$ pour les illettrés, $i=2$ pour le primaire inachevé, $i=3$ pour le primaire achevé, $i=4$ pour le secondaire

inachevé, $i=5$ pour le secondaire achevé, $i=6$ pour le supérieur inachevé, and $i=7$ pour le supérieur achevé.

- $n_{c,i,t}^a$ représente la taille de la population de la tranche d'âge "a" du pays "c" ayant atteint le niveau d'éducation "i" à la date "t".
- $N_{R,t}^G$ représente la taille de la population du grand groupe d'âge "G" appartenant à la région "R" à la date "t". ($N_{R,t}^G = \sum_{c \in R} \sum_{a \in G} \sum_{i=1}^n n_{c,i,t}^a = \sum_{c \in R} N_{c,t}^G = \sum_{c \in R} N_{R,t}^a$).
- $y_{c,i,t}^a$ traduit le nombre d'années de scolarisation cumulées de la tranche d'âge "a" du pays "c" pour atteindre le niveau d'éducation "j" à la date "t".

1.3.2. Calcul des durées des cycles d'enseignement sur la période 1950-2010

Afin de retrouver l'ensemble des durées des cycles d'enseignement utilisées par Barro et Lee (2013), nous avons fusionné les deux bases de la population totale et féminine {MF et F} de Barro et Lee (2013) en utilisant les moyennes d'années de scolarisation par tranches d'âge de chaque cycle. Par la suite nous avons construit un ensemble de 24674 systèmes d'équations à sept inconnues chacun, ce qui nous a permis de retrouver « $y_{c,j,t}^a$ » qui correspond au nombre nécessaire d'années cumulées de scolarisation pour atteindre le niveau d'éducation "j" par la tranche d'âge "a" du pays "c" à la date "t" tel que $j=1..7$, $a=1..13$, $c=1..146$ et $t=1950, 1955..2010$.

Nous nous sommes basés sur les formules des moyennes d'années de scolarisation par cycle. Toutefois, il faut d'abord disposer des proportions d'individus qui n'ont pas achevé les trois cycles d'enseignement. En effet, la base de Barro et Lee (2013) ne les présente pas explicitement. Nous les avons calculées en faisant la différence entre la proportion de ceux qui ont atteint le cycle d'enseignement {notée "lp" "ls" "lh" pour les cycles primaire, secondaire et supérieur respectivement} en question et ceux qui l'ont terminé {notée "lpc" "lsc" "lhc" pour les cycles primaire, secondaire et supérieur respectivement}.

Ainsi, la moyenne d'années de scolarisation du cycle primaire peut être définie comme étant la somme du nombre d'années de scolarisation correspondant aux individus qui n'ont pas achevé le cycle primaire « $y_{c,2,t}^a - y_{c,1,t}^a$ » pondéré par la proportion de ces derniers « $p_{c,2,t}^a$ » plus le nombre d'années de scolarisation correspondant aux individus qui ont achevé le cycle primaire « $y_{c,3,t}^a - y_{c,1,t}^a$ » pondéré par la proportion des individus ayant au moins complété ce même niveau « $\sum_{j=3}^7 p_{c,j,t}^a$ », d'où la formule ci-après :

$$aysp_{c,t}^a = \underbrace{p_{c,2,t}^a [y_{c,2,t}^a - y_{c,1,t}^a]}_{\text{Cycle primaire inachevé}} + \underbrace{\left(\sum_{j=3}^7 p_{c,j,t}^a \right) [y_{c,3,t}^a - y_{c,1,t}^a]}_{\text{Cycle primaire achevé}} \quad (1)$$

Les moyennes d'années de scolarisation des cycles secondaire et supérieur peuvent être obtenues en suivant le même raisonnement:

$$ayss_{c,t}^a = \underbrace{p_{c,4,t}^a [y_{c,4,t}^a - y_{c,3,t}^a]}_{\text{Cycle sec ondaire inachevé}} + \underbrace{\left(\sum_{j=5}^7 p_{c,j,t}^a \right) [y_{c,5,t}^a - y_{c,3,t}^a]}_{\text{Cycle sec ondaire achevé}} \quad (9)$$

$$ayst_{c,t}^a = \underbrace{p_{c,6,t}^a [y_{c,6,t}^a - y_{c,5,t}^a]}_{\text{Cycle supérieur inachevé}} + \underbrace{p_{c,7,t}^a [y_{c,7,t}^a - y_{c,5,t}^a]}_{\text{Cycle supérieur achevé}} \quad (10)$$

L'objectif est de calculer les « $y_{c,j,t}^a$ » pour chaque quadruplet (a,c,t,j) où "a" est la tranche d'âge "c" le pays, "t" est la date et "j" est le niveau d'éducation. Ceci revient à retrouver pour chaque triplet (a,c,t) les sept inconnues $y_{c,1,t}^a; y_{c,2,t}^a; y_{c,3,t}^a; y_{c,4,t}^a; y_{c,5,t}^a; y_{c,6,t}^a; y_{c,7,t}^a$. On note ainsi:

- $aysp_{(mf)_t}^a$: moyenne d'années de scolarisation du cycle primaire de la tranche d'âge « a » de la population totale à la date « t ».
- $ayss_{(mf)_t}^a$: moyenne d'années de scolarisation du cycle secondaire de la tranche d'âge « a » de la population totale à la date « t ».
- $ayst_{(mf)_t}^a$: moyenne d'années de scolarisation du cycle supérieur de la tranche d'âge « a » de la population totale à la date « t ».
- $aysp_{(f)_t}^a$: moyenne d'années de scolarisation du cycle primaire de la tranche d'âge « a » de la population féminine à la date « t ».
- $ayss_{(f)_t}^a$: moyenne d'années de scolarisation du cycle secondaire de la tranche d'âge « a » de la population féminine à la date « t ».
- $ayst_{(f)_t}^a$: moyenne d'années de scolarisation du cycle supérieur de la tranche d'âge « a » de la population féminine à la date « t ».
- $P_{(mf)j,t}^a$: proportion de la population totale appartenant à la tranche d'âge « a » ayant atteint le niveau d'éducation « j » à la date « t ».
- $P_{(f)j,t}^a$: proportion de la population féminine appartenant à la tranche d'âge « a » ayant atteint le niveau d'éducation « j » à la date « t ».

Etant donné que le nombre d'années de scolarisation cumulées $y_{c,j,t}^a$ est le même pour les hommes que pour les femmes $y_{(mf)c,j,t}^a = y_{(m)c,j,t}^a = y_{(f)c,j,t}^a$, nous avons reproduit les trois équations relatives à la moyenne d'années de scolarisation des trois cycles pour la population totale et féminine ($3 \times 2 = 6$ équations). En outre, la dotation éducative de la population non scolarisée est nulle ($y_{c,1,t}^a = 0$). Au total, nous avons pu obtenir pour chaque triplet (a,c,t) un système de sept équations et de sept inconnues.

$$\left\{ \begin{array}{l} aysp_{(mf)_{c,t}}^a = P_{(mf)_{c,2,t}}^a [y_{c,2,t}^a - y_{c,1,t}^a] + \left(\sum_{j=3}^7 P_{(mf)_{c,j,t}}^a \right) [y_{c,3,t}^a - y_{c,1,t}^a] \\ aysp_{(f)_{c,t}}^a = P_{(f)_{c,2,t}}^a [y_{c,2,t}^a - y_{c,1,t}^a] + \left(\sum_{j=3}^7 P_{(f)_{c,j,t}}^a \right) [y_{c,3,t}^a - y_{c,1,t}^a] \\ ayss_{(mf)_{c,t}}^a = P_{(mf)_{c,4,t}}^a [y_{c,4,t}^a - y_{c,3,t}^a] + \left(\sum_{j=5}^7 P_{(mf)_{c,j,t}}^a \right) [y_{c,5,t}^a - y_{c,3,t}^a] \\ ayss_{(f)_{c,t}}^a = P_{(f)_{c,4,t}}^a [y_{c,4,t}^a - y_{c,3,t}^a] + \left(\sum_{j=5}^7 P_{(f)_{c,j,t}}^a \right) [y_{c,5,t}^a - y_{c,3,t}^a] \\ ayst_{(mf)_{c,t}}^a = P_{(mf)_{c,6,t}}^a [y_{c,6,t}^a - y_{c,5,t}^a] + P_{(mf)_{c,7,t}}^a [y_{c,7,t}^a - y_{c,5,t}^a] \\ ayst_{(f)_{c,t}}^a = P_{(f)_{c,6,t}}^a [y_{c,6,t}^a - y_{c,5,t}^a] + P_{(f)_{c,7,t}}^a [y_{c,7,t}^a - y_{c,5,t}^a] \\ y_{c,1,t}^a = 0 \end{array} \right.$$

Ainsi on peut résoudre ce système d'équation pour chaque triplet (a,c,t) et les sept inconnues $\{y_{c,1,t}^a; y_{c,2,t}^a; y_{c,3,t}^a; y_{c,4,t}^a; y_{c,5,t}^a; y_{c,6,t}^a; y_{c,7,t}^a\}$ sont retrouvés comme suit :

$$\begin{cases}
y_{c,1,t}^a = 0 \\
y_{c,2,t}^a = \frac{\left(\sum_{j=3}^7 P(f)_{c,j,t}^a\right) \left(aysp(mf)_{c,t}^a\right) - \left(\sum_{j=3}^7 P(mf)_{c,j,t}^a\right) \left(aysp(f)_{c,t}^a\right)}{P(mf)_{c,2,t}^a \left(\sum_{j=3}^7 P(f)_{c,j,t}^a\right) - P(f)_{c,2,t}^a \left(\sum_{j=3}^7 P(mf)_{c,j,t}^a\right)} \\
y_{c,3,t}^a = \frac{P(mf)_{c,2,t}^a \left(aysp(f)_{c,t}^a\right) - P(f)_{c,2,t}^a \left(aysp(mf)_{c,t}^a\right)}{P(mf)_{c,2,t}^a \left(\sum_{j=3}^7 P(f)_{c,j,t}^a\right) - P(f)_{c,2,t}^a \left(\sum_{j=3}^7 P(mf)_{c,j,t}^a\right)} \\
y_{c,4,t}^a = y_{c,3,t}^a + \frac{\left(\sum_{j=5}^7 P(f)_{c,j,t}^a\right) \left(ayss(mf)_{c,t}^a\right) - \left(\sum_{j=5}^7 P(mf)_{c,j,t}^a\right) \left(ayss(f)_{c,t}^a\right)}{P(mf)_{c,4,t}^a \left(\sum_{j=5}^7 P(f)_{c,j,t}^a\right) - P(f)_{c,4,t}^a \left(\sum_{j=5}^7 P(mf)_{c,j,t}^a\right)} \\
y_{c,5,t}^a = y_{c,3,t}^a + \frac{P(mf)_{c,4,t}^a \left(ayss(f)_{c,t}^a\right) - P(f)_{c,4,t}^a \left(ayss(mf)_{c,t}^a\right)}{P(mf)_{c,4,t}^a \left(\sum_{j=5}^7 P(f)_{c,j,t}^a\right) - P(f)_{c,4,t}^a \left(\sum_{j=5}^7 P(mf)_{c,j,t}^a\right)} \\
y_{c,6,t}^a = y_{c,5,t}^a + \frac{P(f)_{c,7,t}^a \left(ayst(mf)_{c,t}^a\right) - P(mf)_{c,7,t}^a \left(ayst(f)_{c,t}^a\right)}{P(mf)_{c,6,t}^a P(f)_{c,7,t}^a - P(f)_{c,6,t}^a P(mf)_{c,7,t}^a} \\
y_{c,7,t}^a = y_{c,5,t}^a + \frac{P(mf)_{c,6,t}^a \left(ayst(f)_{c,t}^a\right) - P(f)_{c,6,t}^a \left(ayst(mf)_{c,t}^a\right)}{P(mf)_{c,6,t}^a P(f)_{c,7,t}^a - P(f)_{c,6,t}^a P(mf)_{c,7,t}^a}
\end{cases}$$

Il faut préciser que lorsque le dénominateur est nul, on peut retrouver la valeur de $y_{c,j,t}^a$ par l'équation suivante ($y_{c,j,t-5}^{a-1} = y_{c,j,t}^a = y_{c,j,t+5}^{a+1}$), puisqu'il s'agit de la même cohorte.

1.4. Relations entre l'indice de Gini de l'éducation, le taux d'analphabétisme et la moyenne d'années de scolarisation

1.4.1. L'indice de Gini de l'éducation et le taux d'analphabétisme

Afin de mettre en évidence le lien qui existe entre les indices d'inégalités éducatives et le taux d'analphabétisme, nous présentons d'abord la formulation donnée par Morrisson et Murtin (2013). Les auteurs proposent un lien systématique entre d'une part les indices de Gini et Theil de l'éducation et d'autre part le taux d'analphabétisme. Ils considèrent deux distributions f et g d'une même variable d'éducation X telles que la première prenne ses valeurs sur le domaine $[m - M]$ tandis que l'autre, tronquée de sa valeur minimale, se situe dans les valeurs $]m - M]$ où p est la fraction de la population ayant pour valeur de la variable d'éducation m ($x=m$).

Premièrement, les auteurs établissent la relation générale qui lie les indices d'inégalité des deux distributions f et g comme suit :

$$\begin{cases}
Gini(f) = p \times \frac{\mu(f) - m}{\mu(f)} + (1 - p) \times \frac{\mu(f) - mp}{\mu(f)} Gini(g) \\
GE_\alpha(f) = \frac{1}{\alpha(\alpha - 1)} \left[\delta_\alpha(f) + \left(\frac{mp}{\mu(f)}\right)^\alpha - 1 \right] + \delta_\alpha(f) \times GE_\alpha(g)
\end{cases}$$

$$\text{Avec } \delta_\alpha(f) = (1 - p)^{1-\alpha} \times \left(\frac{\mu(f) - mp}{\mu(f)}\right)^\alpha$$

Deuxièmement, Morrisson et Murin (2013) considèrent un cas particulier de la relation générale. Il s'agit du cas où $m=0$ et la variable d'éducation est "le niveau d'éducation maximum atteint". Dans ce cas la relation devient :

$$\begin{cases} Gini(f) = p + (1 - p) \times Gini(g) \\ GE_\alpha(f) = \frac{[(1 - p)^{1-\alpha} - 1]}{\alpha(\alpha - 1)} + (1 - p)^{1-\alpha} \times GE_\alpha(g) \\ Theil(f) = -\ln(1 - p) + Theil(g) \left\{ Theil = \lim_{\alpha \rightarrow 1} GE_\alpha \right\} \end{cases}$$

Ainsi Morrisson et Murin (2013) démontrent que les indices d'inégalité d'éducation sont fortement liés au taux d'analphabétisme. De plus, ils soutiennent que la baisse de l'indice de Gini de l'éducation est expliquée, en grande partie, par la diminution du taux d'analphabétisme.

Par ailleurs, on peut aussi démontrer que le taux d'analphabétisme représente l'inégalité entre les illettrés et les non illettrés (inégalité *between*) par rapport à l'indice de Gini de l'éducation. En effet, si l'on considère deux groupes ($K = 2$) (les illettrés et les non illettrés). Etant donné que $\mu = q_1\mu_1 + q_2\mu_2$ où $q_1 = p_1$ est la proportion des illettrés et $q_2 = (1 - p_1)$ est la proportion des personnes éduquées (non illettrées), la moyenne d'années de scolarisation des personnes éduquées (non illettrées) est égale à: $\mu_2 = \frac{\mu}{q_2} = \frac{\mu}{(1-p_1)}$ puisque la dotation éducative des illettrés est nulle ($\mu_1 = 0$). Ainsi l'inégalité *between* issue de la décomposition de l'indice de Gini par rapport à ces deux groupes peut être calculée comme suit :

$$Gini_B = \left[\frac{1}{2\mu} \sum_{k=1}^K \sum_{k'=1}^K q_k |\mu_k - \mu_{k'}| q_{k'} \right] = \frac{2q_1|\mu_1 - \mu_2|q_2}{2\mu} = \frac{p_1 \left| -\frac{\mu}{(1-p_1)} \right| (1-p_1)}{\mu}$$

$$Gini_B = p_1$$

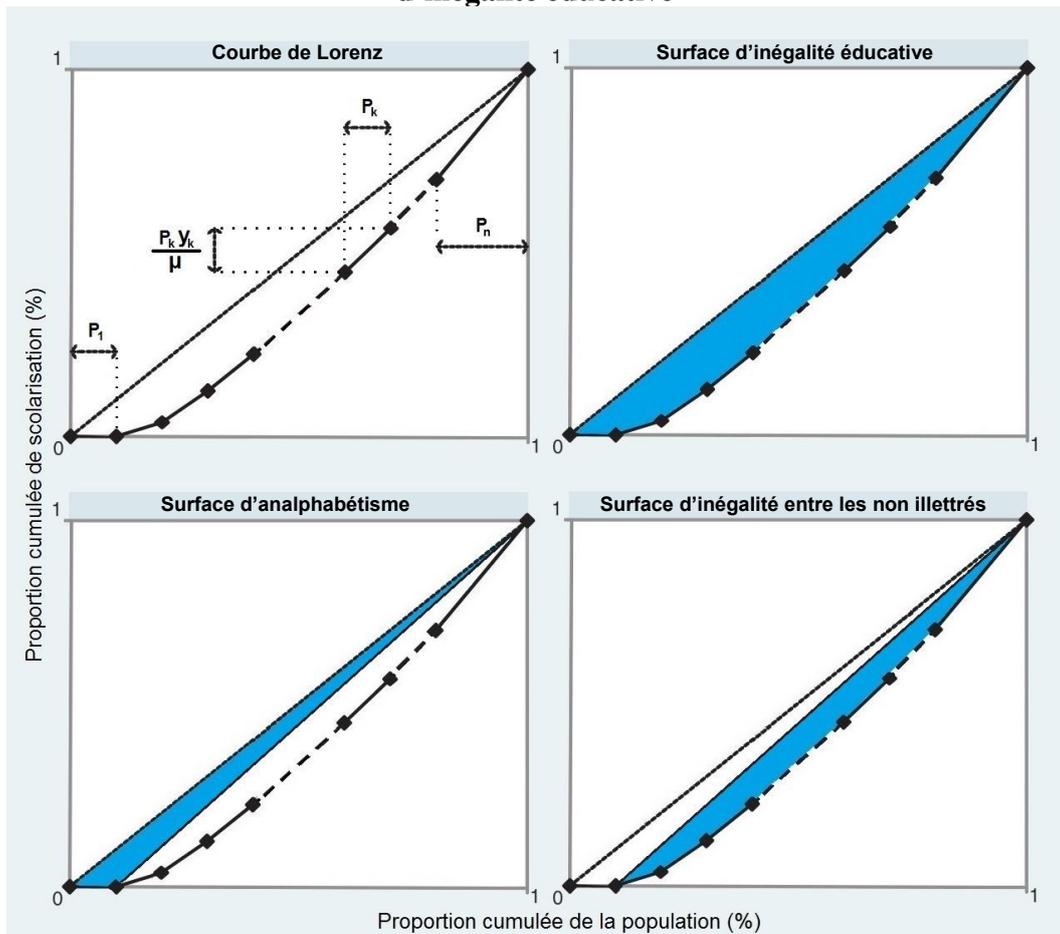
Il faut aussi souligner que dans le cas de cette décomposition, l'indice de Gini est parfaitement décomposable étant donné que la composante *within* correspond exactement à la différence entre l'indice de Gini et la composante *between* ($Gini_W = Gini - Gini_B$) et la contribution des inégalités d'éducation inhérentes à l'intensité de la transvariation est dans ce cas nulle: $Gini_T = 0$.

$$Gini_W = \left[\sum_{k=1}^K \frac{q_k^2 \mu_k}{\mu} Gini_k \right] = \frac{q_1^2 \mu_1}{\mu} Gini_1 + \frac{q_2^2 \mu_2}{\mu} Gini_2 = \frac{(1-p_1)^2 \left[\frac{\mu}{(1-p_1)} \right] Gini_2}{\mu}$$

$$Gini_W = (1-p_1)Gini_2 = Gini - Gini_B$$

D'un point de vue graphique, la courbe de Lorenz nous renseigne aussi sur le lien entre l'indice de Gini de l'éducation et le taux d'analphabétisme. Comme le montre le graphique n°3, la surface d'inégalité éducative en coloré – qui représente la moitié de la valeur de l'indice de Gini de l'éducation – peut être décomposée en deux parties connexes. Une partie composée d'un triangle ayant comme base " p_1 " et de hauteur "1" et dont la surface est égale à la moitié du taux d'analphabétisme et une autre partie liée au reste de la courbe de Lorenz représentant l'inégalité entre les non-illettrés. Ainsi, plus le taux d'analphabétisme est élevé, plus la surface du triangle est importante et plus la distribution de l'éducation est inégalitaire. Autrement dit, l'indice de Gini de l'éducation devient de moins en moins dépendant du taux d'analphabétisme si ce dernier se situe à des niveaux relativement faibles.

Graphique 3 : Contribution graphique de l'analphabétisme à la surface d'inégalité éducative



Source : Auteur.

1.4.2. Le lien entre l'indice de Gini de l'éducation et le niveau moyen d'éducation: vers une formulation générale

Londoño (1990), Ram (1990) et Thomas et al. (2001) ont constaté que la relation entre l'inégalité dans l'éducation, mesurée par l'écart-type de scolarisation, et le niveau de scolarisation suit une courbe en cloche. Lorsqu'un pays part d'un niveau d'instruction très faible pour parvenir à un niveau maximum, la variance commence par augmenter avant de régresser. Les résultats des travaux empiriques indiquent que la dispersion du nombre d'années d'étude augmente de manière continue à mesure qu'augmente le niveau moyen de scolarisation jusqu'à un certain pic – qui correspond à environ 7 années de scolarisation – puis diminue. Toutefois, Thomas et al. (2001) constatent qu'il existe une corrélation négative très forte entre l'inégalité dans l'éducation, mesurée cette fois-ci par l'indice de Gini de l'éducation et le nombre moyen d'années d'étude pour 85 pays sur plusieurs décennies.

Les estimations en coupe transversale ou en panel montrent de manière équivalente une relation négative très significative. Partant de l'analyse empirique, beaucoup d'auteurs soulignent que la relation entre inégalités dans l'éducation et nombre moyen d'années de scolarisation n'est pas linéaire : l'inégalité d'éducation est toujours plus forte lorsque la durée moyenne de scolarisation est faible et s'atténue au fil de son évolution (Ram, 1990 ; Londoño, 1990 ; De Gregorio et Lee, 2002 ; Lim et Tang, 2008 ; Morrisson et Murin, 2013).

En se basant aussi sur la courbe de Lorenz, Berthélemy (2006) suggère une reformulation de l'indice de Gini de l'éducation qui permet, selon lui, d'extraire la moyenne d'années de scolarisation de l'indice de Gini de l'éducation. En effet, sous certaines hypothèses simplificatrices, l'auteur propose une méthodologie permettant de séparer la composante propre au niveau d'éducation de celle relative à la distribution de la scolarisation dans la formule de l'indice de Gini de l'éducation.

La première hypothèse concerne le nombre de niveaux d'éducation retenus. L'auteur considère uniquement quatre niveaux d'éducation ($n=4$): aucune scolarisation (pour les illettrés), les niveaux primaire, secondaire et supérieur. La deuxième hypothèse consiste à supposer que les trois cycles d'enseignement primaire, secondaire et supérieur, ont des durées d'enseignement de taille égale.

En plus de ces deux hypothèses, Berthélemy (2006) adopte implicitement une troisième hypothèse simplificatrice que nous n'utilisons pas ici : la durée de scolarisation de tous les cycles d'enseignement (primaire, secondaire et supérieur) est unitaire. Ainsi l'auteur suppose que :

$$\begin{cases} y_1 = 0 \\ y_2 = 1 \\ y_3 = 2 \\ y_4 = 3 \end{cases}$$

• **Cas de quatre niveaux d'éducation (Berthélemy, 2006)**

Dans un premier temps, en retenant uniquement les deux premières hypothèses utilisées par Berthélemy (2006), nous considérons contrairement à l'auteur que la durée de scolarisation du cycle primaire est égale à "y" (où y n'est pas forcément égale à 1). Ainsi on a :

$$\begin{cases} y_1 = 0 \\ y_2 = y \\ y_3 = 2y \\ y_4 = 3y \end{cases}$$

Tout comme Berthélemy (2006), au lieu d'utiliser les proportions des individus ayant respectivement les quatre niveaux retenus²⁶, nous proposons de réécrire la moyenne d'années de scolarisation et l'indice de Gini de l'éducation en fonction des proportions des individus ayant respectivement *au moins* les niveaux d'enseignement primaire, secondaire et supérieur notés (s_1 , s_2 et s_3) de telle façon que :

$$\begin{cases} p_1 = 1 - s_1 \\ p_2 = s_1 - s_2 \\ p_3 = s_2 - s_3 \\ p_4 = s_3 \end{cases}$$

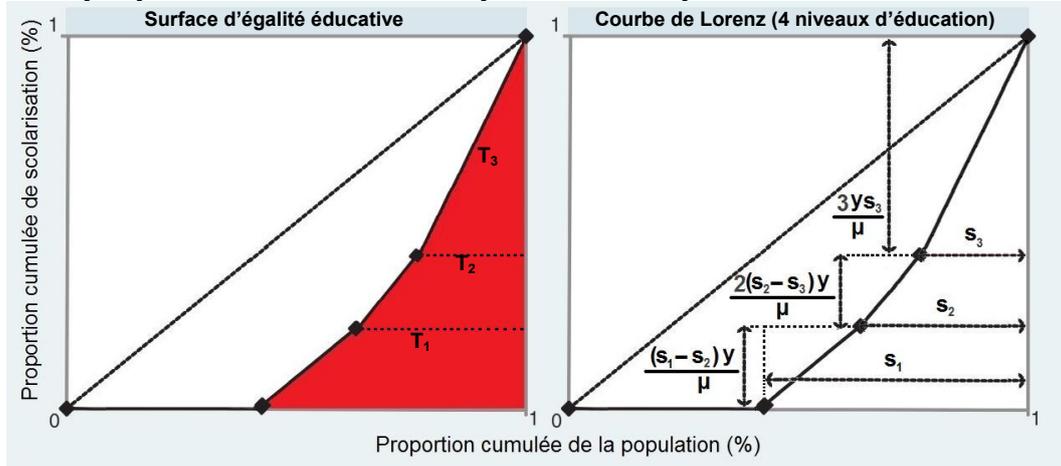
Ainsi, la moyenne d'années de scolarisation peut être définie en fonction de s_1 , s_2 et s_3 comme suit:

²⁶ Les proportions des individus p_1 , p_2 , p_3 et p_4 sont retenues respectivement pour les niveaux suivants : aucune scolarisation, primaire, secondaire et supérieur, avec $p_1 + p_2 + p_3 + p_4 = 1$.

$$\mu = \sum_{j=1}^4 y_j p_j = 0 \times (1 - s_1) + y \times (s_1 - s_2) + 2y \times (s_2 - s_3) + 3y \times (s_3)$$

$$\mu = y \times (s_1 + s_2 + s_3)$$

Graphique 4 : Courbe de Lorenz pour le cas de quatre niveaux d'éducation



Source : Auteur.

De même, tout comme Berthélemy (2006), nous proposons une formulation de l'indice de Gini de l'éducation en fonction de s_1 , s_2 et s_3 . Nous nous basons sur la surface colorée dans le graphique n°4 (surface d'égalité de l'éducation) que nous notons "S". Si l'indice de Gini de l'éducation représente le double de la surface comprise entre la diagonale et la courbe de Lorenz (surface d'inégalité), nous pouvons aussi le définir par rapport à la surface d'égalité de l'éducation en utilisant la formule suivante : $Gini = 1 - 2S$. Cette dernière correspond à la surface de deux trapèzes et d'un triangle (respectivement T_1 , T_2 et T_3) comme le montre le graphique n° 4.

Ainsi on a alors :

$$\begin{cases} T_1 = \frac{1}{2} (s_1 + s_2) \times \left(\frac{(s_1 - s_2)y}{\mu} \right) = \frac{y}{2\mu} \times (s_1^2 - s_2^2) \\ T_2 = \frac{1}{2} (s_2 + s_3) \times \left(\frac{2(s_2 - s_3)y}{\mu} \right) = \frac{y}{2\mu} \times (2s_2^2 - 2s_3^2) \\ T_3 = \frac{1}{2} s_3 \times \left(\frac{3ys_3}{\mu} \right) = \frac{y}{2\mu} \times (3s_3^2) \end{cases}$$

Etant donné que $\mu = y \times (s_1 + s_2 + s_3)$, la surface "S" s'écrit comme suit :

$$S = T_1 + T_2 + T_3 = \frac{y}{2\mu} \times ((s_1^2 - s_2^2) + 2(s_2^2 - s_3^2) + 3s_3^2) = \frac{(s_1^2 + s_2^2 + s_3^2)}{2(s_1 + s_2 + s_3)}$$

D'où l'indice de Gini de l'éducation est égal à :

$$Gini = 1 - 2S = 1 - \frac{(s_1^2 + s_2^2 + s_3^2)}{(s_1 + s_2 + s_3)}$$

Subséquentement, sous les deux hypothèses retenues, les formulations de la moyenne d'années de scolarisation de l'indice de Gini de l'éducation sont :

$$\begin{cases} \mu = y \times (s_1 + s_2 + s_3) \\ Gini = 1 - \frac{(s_1^2 + s_2^2 + s_3^2)}{(s_1 + s_2 + s_3)} \end{cases}$$

Par ailleurs, Berthélemy (2006) – qui suppose de plus que $y = 1$ – propose de réécrire l'indice de Gini de manière à faire apparaître la moyenne d'années de scolarisation μ :

$$Gini = 1 - 2S = 1 - \mu \left[1 - 2 \frac{(\sigma + \sigma\tau + \tau)}{(1 + \sigma + \tau)^2} \right] \text{ où } \begin{cases} \sigma = s_2/s_1 \\ \tau = s_3/s_1 \end{cases}$$

L'auteur conclut par la suite que l'on peut séparer la moyenne d'années de scolarisation de l'indice de Gini de l'éducation puisqu'il retient que le terme $\left[1 - 2 \frac{(\sigma + \sigma\tau + \tau)}{(1 + \sigma + \tau)^2} \right]$ dépend uniquement de la structure des taux de scolarisation du système éducatif et non pas de la moyenne d'années de scolarisation.

Toutefois, Berthélemy (2006) considère que les termes σ et τ correspondent respectivement au rapport entre le taux de scolarisation du secondaire et du primaire et au rapport entre le taux de scolarisation du supérieur et du primaire. Néanmoins, l'expression de chacun de ces deux termes se réfère à la répartition de toute la population en question et non pas uniquement à celle appartenant au système éducatif. Par conséquent, ils n'ont pas un lien direct avec les taux de scolarisation²⁷. Ainsi, le terme σ représente plutôt la proportion de la population ayant au moins le niveau secondaire rapportée à la proportion de la population ayant au moins le niveau primaire.

Ceci étant, dans le cas où l'on ne maintient pas la troisième hypothèse ($y \neq 1$) la formule de l'indice de Gini de l'éducation proposée par Berthélemy (2006) devient : $Gini = 1 - \frac{\mu}{y} \left[1 - 2 \frac{(\sigma + \sigma\tau + \tau)}{(1 + \sigma + \tau)^2} \right]$. Ainsi il se peut qu'il y ait une éventuelle dépendance entre $\frac{\mu}{y}$ et l'expression entre crochets.

• *Généralisation dans le cas de « n » niveaux d'éducation*

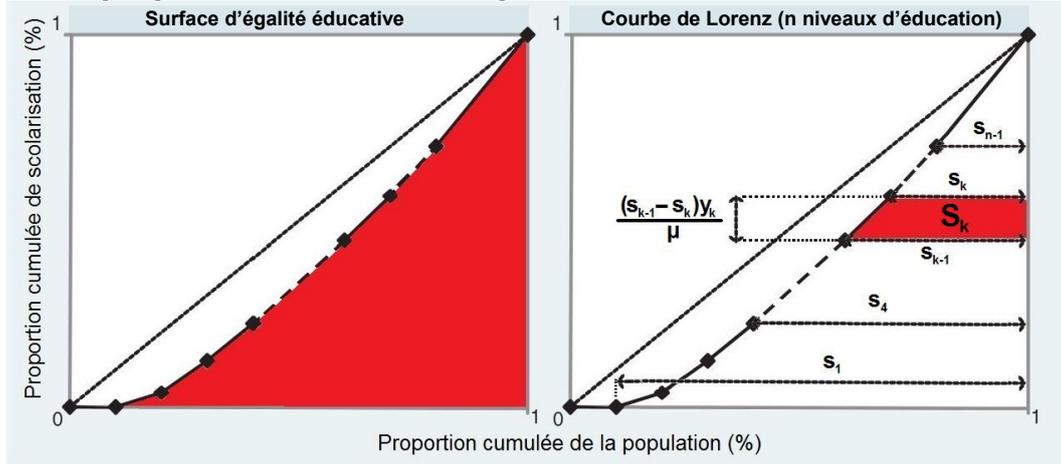
Dans un deuxième temps, nous reprenons la formulation de l'indice de Gini de l'éducation en se plaçant cette fois-ci dans un cadre plus général. Ainsi nous ignorons dans ce qui suit les trois hypothèses retenues par Berthélemy (2006).

Nous retenons ainsi "n" niveaux d'éducation où s_{k-1} est défini comme étant la proportion des individus ayant au moins y_k années de scolarisation. On note $s_0 = 1$ et $s_n = 0$. Etant donné que $p_k = (s_{k-1} - s_k)$ et $y_1 = 0$, la moyenne d'années de scolarisation peut se réécrire comme suit :

$$\begin{aligned} \mu &= \sum_{k=1}^n y_k p_k = \sum_{k=1}^n y_k (s_{k-1} - s_k) = y_1 s_0 - y_n s_n + \sum_{k=1}^{n-1} s_k (y_{k+1} - y_k) \\ \mu &= \sum_{k=1}^{n-1} s_k (y_{k+1} - y_k) \end{aligned}$$

²⁷A titre d'exemple, le calcul du taux de scolarisation en primaire s'appuie uniquement sur l'effectif des élèves en cours de scolarisation au cycle primaire et non pas sur l'effectif des individus dans la population ayant le niveau primaire.

Graphique 5 : Courbe de Lorenz pour le cas de "n" niveaux d'éducation



Source : Auteur.

La surface S correspond à la somme des surfaces S_k du graphique n° 5 :

$$S_k = \frac{1}{2}(s_{k-1} + s_k) \times \left(\frac{(s_{k-1} - s_k)y_k}{\mu} \right) = \frac{1}{2} \left(\frac{(s_{k-1}^2 - s_k^2)y_k}{\mu} \right)$$

Ainsi puisque

$$2S = \sum_{k=1}^n 2S_k = \frac{1}{\mu} \sum_{k=1}^n (s_{k-1}^2 - s_k^2)y_k = \frac{1}{\mu} \left[\sum_{k=1}^{n-1} s_k^2 (y_{k+1} - y_k) \right]$$

l'indice de Gini de l'éducation peut être déduit comme suit :

$$Gini = 1 - 2S = 1 - \left[\sum_{k=1}^{n-1} s_k^2 (y_{k+1} - y_k) \right] / \left[\sum_{k=1}^{n-1} s_k (y_{k+1} - y_k) \right]$$

Etant donné que y_n correspond à la durée de scolarisation maximale²⁸. On définit l'indice de scolarisation (I_μ) comme étant le rapport de la moyenne d'années de scolarisation sur sa durée théorique maximale ($I_\mu = \frac{\mu}{y_n}$) qui est un indice compris entre 0 et 1. On définit aussi le poids de la durée de scolarisation du niveau $k+1$ (w_k) comme étant le rapport de la durée de scolarisation propre au niveau $k+1$ sur la durée théorique maximale y_n ($w_k = d_{k+1}/y_n = [y_{k+1} - y_k]/y_n$). On a ainsi :

$$\sum_{k=1}^{n-1} w_k = \sum_{k=1}^{n-1} [y_{k+1} - y_k]/y_n = \frac{[y_n - y_1]}{y_n} = 1$$

Si on note respectivement $M_a(s_k, w_k)$ et $M_q(s_k, w_k)$ les moyennes arithmétique et quadratique des s_k pondérées par les w_k , on peut donc reformuler l'indice de scolarisation et l'indice de Gini de l'éducation comme suit :

²⁸ La durée de scolarisation maximale est la durée cumulée de scolarisation maximale offerte par un système éducatif donné et qui correspond à la somme totale des durées de scolarisation des trois cycles d'enseignement primaire, secondaire et supérieur du système éducatif.

$$\begin{cases} I_\mu = \frac{\mu}{y_n} = \sum_{k=1}^{n-1} w_k s_k = M_a(s_k, w_k) \\ Gini = 1 - I_\mu \times \left(\frac{\left[\sum_{k=1}^{n-1} w_k s_k^2 \right]}{\left[\sum_{k=1}^{n-1} w_k s_k \right]^2} \right) = 1 - I_\mu \times \left[\frac{M_q(s_k, w_k)}{M_a(s_k, w_k)} \right]^2 \end{cases}$$

Dans le cas où I_μ tend vers 1 ($I_\mu \rightarrow 1$), la moyenne quadratique tend vers la moyenne arithmétique ($M_q(s_k, w_k) \rightarrow M_a(s_k, w_k)$) et le rapport $\left[\frac{M_q(s_k, w_k)}{M_a(s_k, w_k)} \right]^2$ devient de plus en plus dépendant de I_μ . Ainsi, comme il n'est pas possible d'extraire la moyenne arithmétique d'une moyenne quadratique, on peut finalement conclure qu'il n'est pas possible de séparer la composante propre au niveau d'éducation de celle relative à la distribution de la scolarisation dans la formule de l'indice de Gini de l'éducation.

Gille (2015) qui a adopté la méthodologie de Berthélemy (2006) afin de corriger l'indice de Gini de l'éducation, rapporte que la correction proposée n'est pas parfaite et qu'il reste encore une corrélation entre la moyenne d'années de scolarisation et l'indice de Gini corrigé. Ce qui corrobore notre conclusion. On ne peut pas dissocier l'indice de Gini de l'éducation de la moyenne d'années de scolarisation.

1.4.3. *Le domaine de variation de l'indice de Gini de l'éducation selon la moyenne d'années de scolarisation*

Etant donné qu'on ne peut extraire mathématiquement l'indice de Gini de l'éducation de la moyenne d'années de scolarisation, nous cherchons à définir l'intervalle dans lequel les valeurs de l'indice de Gini de l'éducation peuvent se situer à chacune des valeurs de la moyenne d'années de scolarisation.

❖ **Borne maximale de l'indice de Gini de l'éducation**

Puisque la moyenne quadratique est toujours supérieure à la moyenne arithmétique on peut donc majorer l'indice de Gini de l'éducation comme suit :

$$M_q(s_k, w_k) \geq M_a(s_k, w_k) \Rightarrow 1 - \frac{M_q(s_k, w_k)^2}{M_a(s_k, w_k)} \leq 1 - M_a(s_k, w_k) \Rightarrow Gini \leq 1 - \frac{\mu}{y_n}$$

Si l'on suppose que $s_k = s, \forall k \in \llbracket 1, n-1 \rrbracket$ on peut déduire que la moyenne d'années de scolarisation est égale à $\mu = y_n \sum_{k=1}^{n-1} w_k s_k = y_n \times s$ et que l'indice de Gini de l'éducation est : $Gini = 1 - s = 1 - \frac{\mu}{y_n}$. Ainsi on déduit que ce majorant correspond à la valeur maximale.

$$\begin{cases} Gini \leq f_{max}(\mu) \\ f_{max}(\mu) = 1 - \frac{\mu}{y_n} \end{cases}$$

❖ **Borne minimale de l'indice de Gini de l'éducation**

Etant donné une moyenne d'années de scolarisation $\mu = y_n \sum_{k=1}^{n-1} w_k s_k$ telle que $y_i < \mu < y_{i+1}$:

1) On peut démontrer que $0 < s_i < 1$ ($s_i \neq 0$: et $s_i \neq 1$)

En suivant un raisonnement par l'absurde on peut démontrer ceci comme suit :

- Si l'on suppose que $s_i = 1$ et puisque $0 \leq s_k \leq s_{k-1} \leq 1, \forall k \in \llbracket 2 - n \rrbracket$ on a alors $\forall k \leq i, s_k = 1$ d'où $\mu \geq y_n \sum_{k=1}^i w_k$ ainsi $\mu \geq y_{i+1}$ ce qui est absurde donc $s_i \neq 1$.
- Si l'on suppose que $s_i = 0$ et puisque $0 \leq s_k \leq s_{k-1} \leq 1, \forall k \in \llbracket 2 - n \rrbracket$ on a alors $\forall k \geq i, s_k = 0$ d'où $\mu \leq y_n \sum_{k=1}^{i-1} w_k$ ainsi $\mu \leq y_i$ ce qui est absurde donc $s_i \neq 0$.

D'où le résultat.

2) On peut démontrer aussi que si $y_i < \mu < y_{i+1}$

$$Z^{Max}(s_1, \dots, s_{n-1}) = y_i + \frac{[\mu - y_i]^2}{y_{i+1} - y_i}$$

si on note :

$$Z(s_1, \dots, s_{n-1}) = y_n [M_q(s_k, w_k)]^2 = y_n \left[\sum_{k=1}^{n-1} w_k s_k^2 \right]$$

Puisqu'on a précédemment démontré que $0 < s_i < 1$ ($s_i^2 \neq s_i$) on a alors $s_k^2 \leq s_k, \forall k \neq i$ donc la fonction $Z(s_1, \dots, s_{n-1})$ est maximale si $\forall k \neq i, s_k^2 = s_k$ ce qui implique que cette dernière est maximale si $\forall k \neq i, s_k(1 - s_k) = 0 \Leftrightarrow$ si ($s_k = 1$ ou $s_k = 0$), Ainsi si $k < i$ on a $s_k \neq 0$ puisque $0 < s_i \leq s_k$ d'où $s_k = 1$ et de même si $k > i$ on a $s_k \neq 1$ puisque $s_k \leq s_i < 1$ d'où $s_k = 0$.

D'où la fonction $Z(s_1, \dots, s_{n-1})$ est maximale si :

$$s_k = \begin{cases} 1 & \text{si } k < i \\ 0 & \text{si } k > i \end{cases}$$

Dans ce cas on a :

$$\begin{aligned} \mu &= y_n \sum_{k=1}^{n-1} w_k s_k = y_n \left[\sum_{k=1}^{i-1} w_k \right] + y_n w_i s_i = y_i + [y_{i+1} - y_i] s_i \\ \Rightarrow s_i &= \left[\frac{\mu - y_i}{y_{i+1} - y_i} \right] \end{aligned}$$

Ainsi :

$$Z^{Max}(s_1, \dots, s_{n-1}) = y_n \left[\sum_{k=1}^{i-1} w_k \right] + y_n w_i s_i^2 = y_i + \frac{[\mu - y_i]^2}{y_{i+1} - y_i}$$

D'où le résultat.

3) On peut enfin déduire que

$$Gini \geq \frac{(\mu - y_i)(y_{i+1} - \mu)}{\mu(y_{i+1} - y_i)} \text{ si } y_i < \mu < y_{i+1}$$

Comme on a démontré précédemment que :

$$Z(s_1, \dots, s_{n-1}) \leq y_i + \frac{[\mu - y_i]^2}{y_{i+1} - y_i}$$

$$\Rightarrow Gini = \frac{\mu - Z(s_1, \dots, s_{n-1})}{\mu} \geq \frac{\mu(y_{i+1} - y_i) - y_i(y_{i+1} - y_i) - [\mu - y_i]^2}{\mu(y_{i+1} - y_i)}$$

$$\Rightarrow Gini \geq \frac{(\mu - y_i)(y_{i+1} - \mu)}{\mu(y_{i+1} - y_i)}$$

Si l'on pose $s_k = \begin{cases} 1 & \text{si } k < i \\ (\mu - y_i)/(y_n w_i) & \text{si } k = i \\ 0 & \text{si } k > i \end{cases}$, de telle sorte que la moyenne

d'années de scolarisation μ vérifie $y_i < \mu < y_{i+1}$, on peut démontrer dans ce cas que l'indice de Gini de l'éducation est égal à $Gini = \frac{(\mu - y_i)(y_{i+1} - \mu)}{\mu(y_{i+1} - y_i)}$ (voir la démonstration dans l'annexe n°3). Cette valeur de *Gini* correspond à la valeur minimale de l'indice de Gini de l'éducation.

$$\begin{cases} Gini \geq f_{min}(\mu) \\ f_{min}(\mu) = \frac{(\mu - y_i)(y_{i+1} - \mu)}{\mu(y_{i+1} - y_i)} \end{cases}$$

Etant donné une valeur de la moyenne d'années de scolarisation μ , on peut définir le domaine dans lequel les valeurs de l'indice de Gini de l'éducation se situent :

$$\begin{cases} f_{min}(\mu) \leq Gini \leq f_{max}(\mu) \\ f_{min}(\mu) = \frac{(\mu - y_i)(y_{i+1} - \mu)}{\mu(y_{i+1} - y_i)} \\ f_{max}(\mu) = 1 - \frac{\mu}{y_n} \end{cases}$$

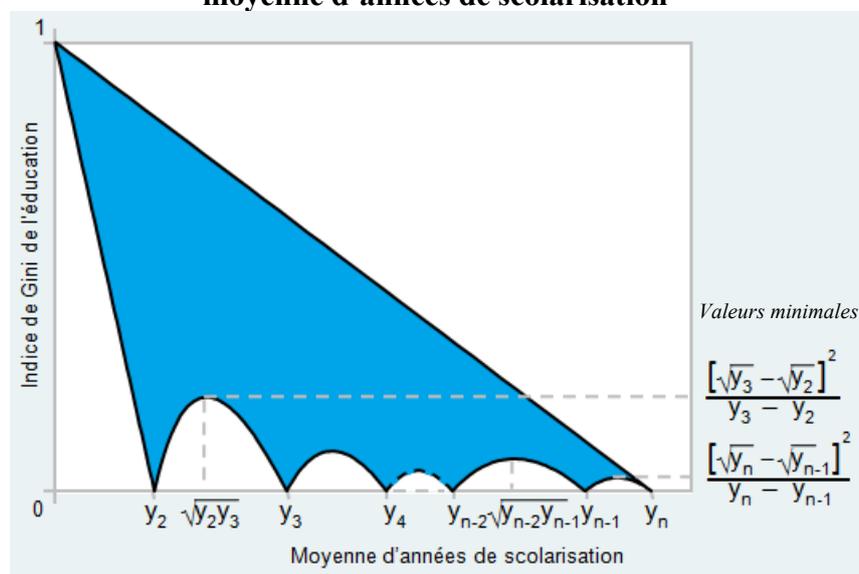
Dans le cas où $y_i \leq \mu \leq y_{i+1}$ tel que $i \neq 1$ la fonction $f_{min}(\mu)$ a la forme d'une courbe en cloche. Elle est nulle pour les deux valeurs $\mu = y_i$ et $\mu = y_{i+1}$ et atteint son maximum $\left(\frac{[\sqrt{y_{i+1}} - \sqrt{y_i}]^2}{|y_{i+1} - y_i|} \right)$ pour la seule valeur $\mu = \sqrt{y_i y_{i+1}}$. Dans le cas où $i = 1$ la fonction $f_{min}(\mu)$ devient une droite passant par les points (0,1) et $(y_2, 0)$ puisque $y_1 = 0$. La fonction $f_{max}(\mu)$, quant à elle, est toujours une droite passant par les points (0,1) et $(y_n, 0)$. On peut ainsi tracer le domaine de variation possible de la moyenne d'années de scolarisation et de l'indice de Gini de l'éducation (graphique n°6).

Comme le montre le graphique n°6, on peut déterminer, pour un système éducatif donné, les valeurs minimale et maximale de l'indice de Gini dès lors que l'on connaît la moyenne d'années de scolarisation (μ) et les durées cumulées d'enseignement (y_i).

En conclusion, même si l'indice de Gini de l'éducation est corrélé à la moyenne d'années de scolarisation, il n'est pas prédéterminé systématiquement par cette dernière. Nous avons montré les limites de la méthodologie suggérée par Berthélemy (2006) qui consiste à extraire de l'indice de Gini la composante relative au niveau de l'éducation. La valeur de l'indice de Gini de l'éducation dépend non seulement de la durée maximale offerte par le système éducatif en question (y_n) mais aussi de l'ensemble des durées cumulées de scolarisation (y_i) utilisées dans

son calcul. Nous avons défini l'intervalle de variation de l'indice de Gini de l'éducation qui peut être associé à la valeur de la moyenne d'années de scolarisation en fonction des durées cumulées d'enseignement (y_i).

Graphique 6 : Domaine de variation de l'indice de Gini de l'éducation selon la moyenne d'années de scolarisation



Source : Auteur.

1.5. Tendances des inégalités éducatives selon les niveaux de développement des pays

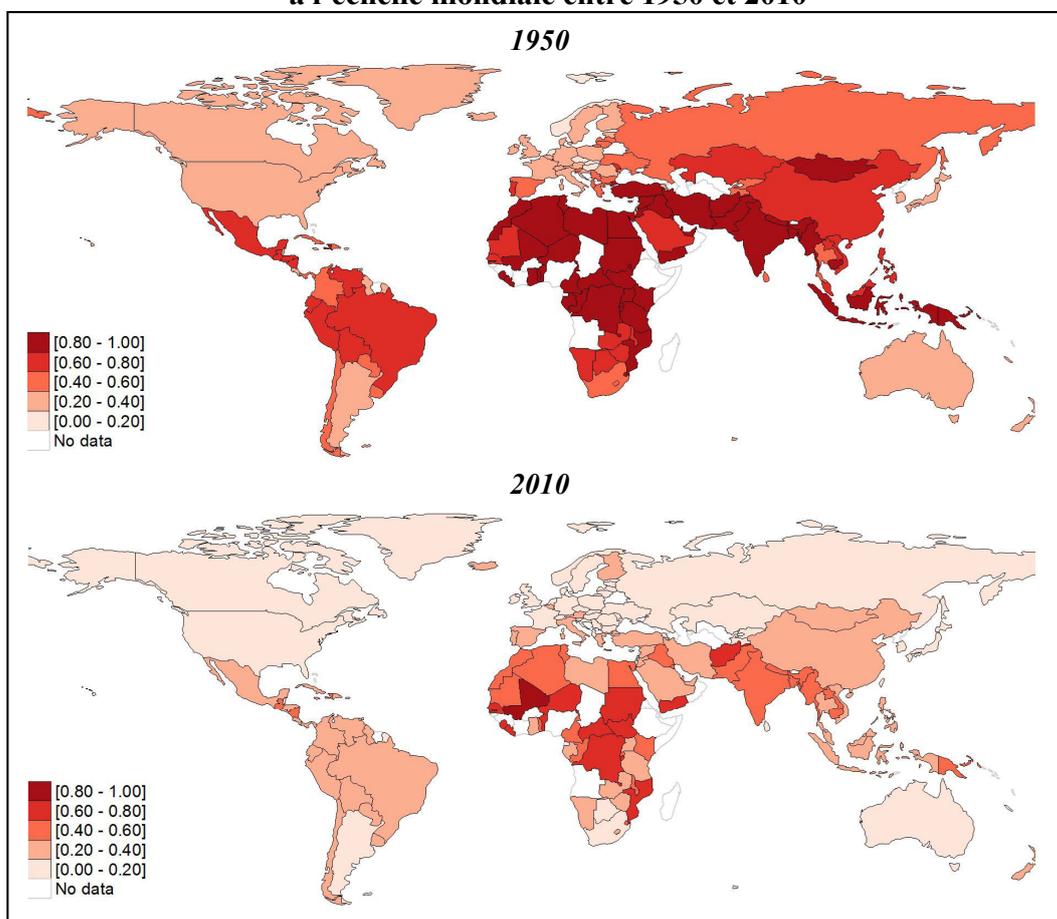
La figure n° 2 présente une carte de l'indice de Gini de l'éducation à travers le monde en 1950 et en 2010. Il apparaît que, durant ces soixante années, d'importants efforts ont été entrepris en termes d'égalisation de l'accès à l'éducation. Les pays ont vu pour la plupart leur indice de Gini de l'éducation baisser de manière remarquable. L'indice de Gini de l'éducation de la population mondiale âgée de 15 et plus a presque été réduit de moitié, il est passé de 0.64 en 1950 à 0.34 en 2010.

Ce déclin est perceptible également si on considère les pays selon le niveau de développement. Pour les pays développés, l'indice de Gini de l'éducation de la population âgée de 15 ans et plus a diminué de 0.38 à 0.19, celui des pays en développement est passé de 0.73 à 0.36 (Tableau n°9). Néanmoins, malgré cette réduction significative des inégalités éducatives, l'écart entre pays développés et en développement demeure assez important. La différence en termes d'efforts alloués à la réduction des inégalités éducatives est de l'ordre de 6 décennies puisque le niveau d'inégalité en 2010 dans les pays en développement est comparable au niveau d'inégalité constaté en 1950 dans les pays développés (Tableau n° 9).

Le tableau n°9 présente également l'indice de Gini mondial de l'éducation par groupes d'âge. S'il y a une baisse pour l'ensemble des groupes d'âge, la réduction de l'inégalité a été particulièrement importante chez les jeunes âgés de 15 à 24 ans. La population âgée de 75 ans et plus constitue le groupe d'âge pour lequel la réduction des inégalités éducatives a été la plus faible : l'indice de Gini mondial de l'éducation a connu une réduction de 57% (de 0.56 en 1950 à 0.24 en 2010) pour

le groupe d'âge 15-24 ans, et de 16% (de 0.61 en 1950 à 0.51 en 2010) pour les 75 ans et plus.

Figure 2 : Indice de Gini de l'éducation de la population âgée de 15 ans et plus à l'échelle mondiale entre 1950 et 2010



Source : Auteur (Base mondiale des inégalités éducatives).

Tableau 9: Indice de Gini de l'éducation par niveau de développement et par groupes d'âge, 1950-2010

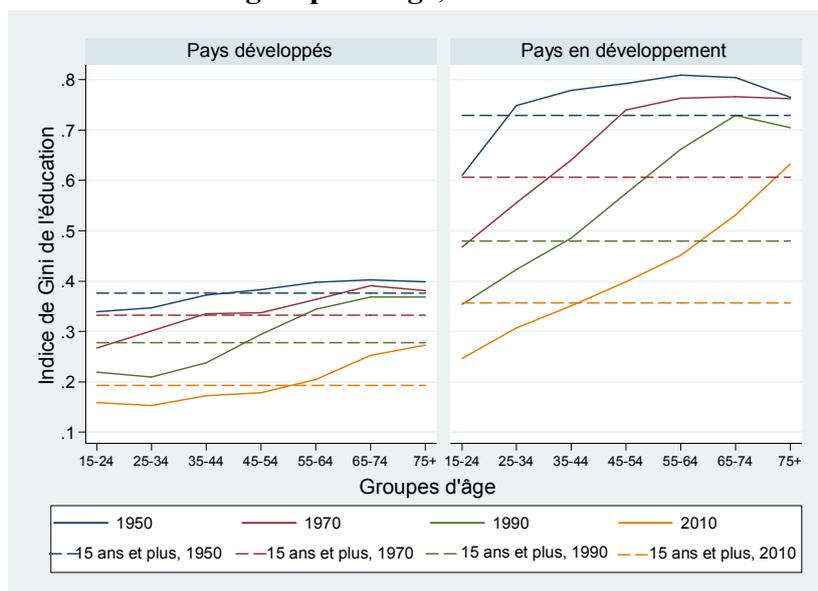
	1950	1970	1990	2010
Niveau de développement				
Pays développés (24)	0.38	0.33	0.28	0.19
Pays en développement (122)	0.73	0.61	0.48	0.36
Groupe d'âge				
15-24	0.56	0.44	0.34	0.24
25-34	0.66	0.51	0.40	0.29
35-44	0.67	0.58	0.45	0.33
45-54	0.67	0.64	0.52	0.37
55-64	0.68	0.65	0.59	0.41
65-74	0.66	0.64	0.63	0.47
75 ans et plus	0.61	0.61	0.57	0.51
Pays du monde (146)	0.64	0.55	0.45	0.34

Source: Auteur (Base mondiale des inégalités éducatives)

Le graphique n°7 présente l'indice de Gini de l'éducation à la fois par niveau de développement et par groupes d'âge. Ainsi, pour la population âgée de plus de 75 ans, l'indice de Gini de l'éducation en 2010 est pour les pays en développement de l'ordre de 0.63 tandis que celui enregistré en 1950 pour les pays développés est

de l'ordre de 0.40. Par conséquent le fossé en termes d'années de décalage pour cette tranche d'âge pourrait dépasser 100 ans.

Graphique 7 : Indice de Gini de l'éducation par niveau de développement et groupes d'âge, 1950-2010



Source: Auteur (Base mondiale des inégalités éducatives)

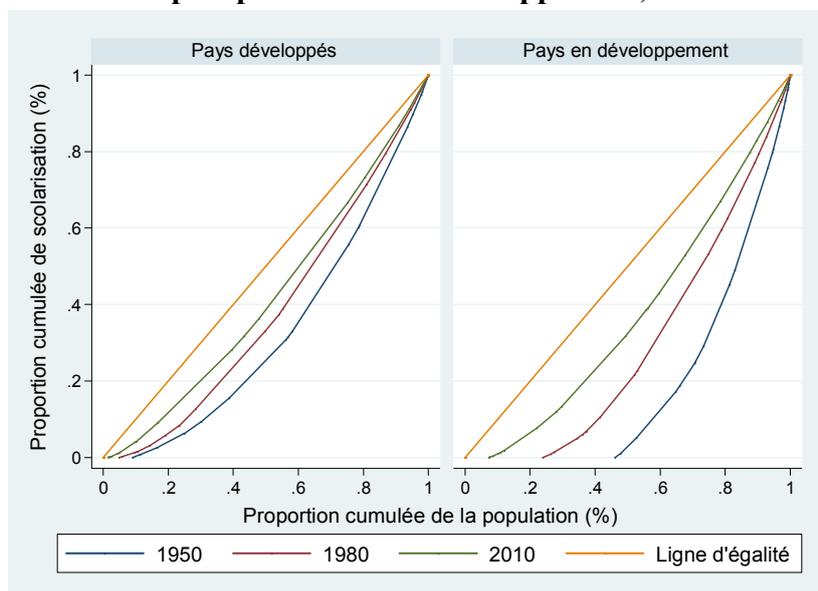
La différence en termes d'années de décalage entre l'indice de Gini de l'éducation des pays développés et des pays en développement est, pour les groupes d'âge [35-44] et [25-34], respectivement de l'ordre de 60 et 40 ans (graphique n°7). Toutefois, cette différence s'est nettement réduite pour atteindre 20 ans pour la population des 15-24 ans (graphique n°7). En effet, le niveau d'inégalité pour cette tranche d'âge en 2010 pour les pays en développement correspond à celui enregistré en 1990 pour les pays développés. A cet égard, il faut souligner que c'est la population des jeunes qui a le plus bénéficié de la progression liée au développement de l'éducation de base durant ces vingt dernières années.

Par ailleurs, le graphique n°8 présente la dynamique de la courbe de Lorenz de l'éducation de la population âgée de 15 ans et plus par niveau de développement entre 1950 et 2010. L'examen de ces courbes, à la fois pour les pays développés et en développement, montre une amélioration de la distribution de l'éducation qui se traduit par un déplacement de ces courbes au cours du temps vers la bissectrice qui représente la ligne égalitaire. Toutefois, la dynamique est différente selon le niveau de développement. La courbe de Lorenz des pays en développement se déplace au cours du temps à travers la partie inférieure de la distribution de l'éducation tandis que celle des pays développés se déplace plutôt à travers la partie supérieure de la distribution. Ainsi, pour les pays développés, la courbe de Lorenz de l'éducation est pratiquement reliée à l'origine du graphique, et ce depuis 1950. Ces pays connaissent un niveau d'analphabétisme réduit dès le début des années 1950, la proportion de la population âgée de 15 et plus n'ayant aucune scolarisation était seulement de 9.2%. Il est passé à 2.4% en 2010.

Dans les pays développés, il semblerait que l'amélioration de l'accès à l'éducation soit passé par l'augmentation de la proportion de la population âgée de 15 et plus ayant atteint le niveau supérieur. Cette dernière est passée de 5.7% en

1950 à 32.2% en 2010. Tandis que les pays en développement ont beaucoup plus élargi l'éducation de base (la partie inférieure de la distribution de l'éducation).

Graphique 8 : La courbe de Lorenz de l'éducation de la population âgée de 15 ans et plus par niveau de développement, 1950-2010

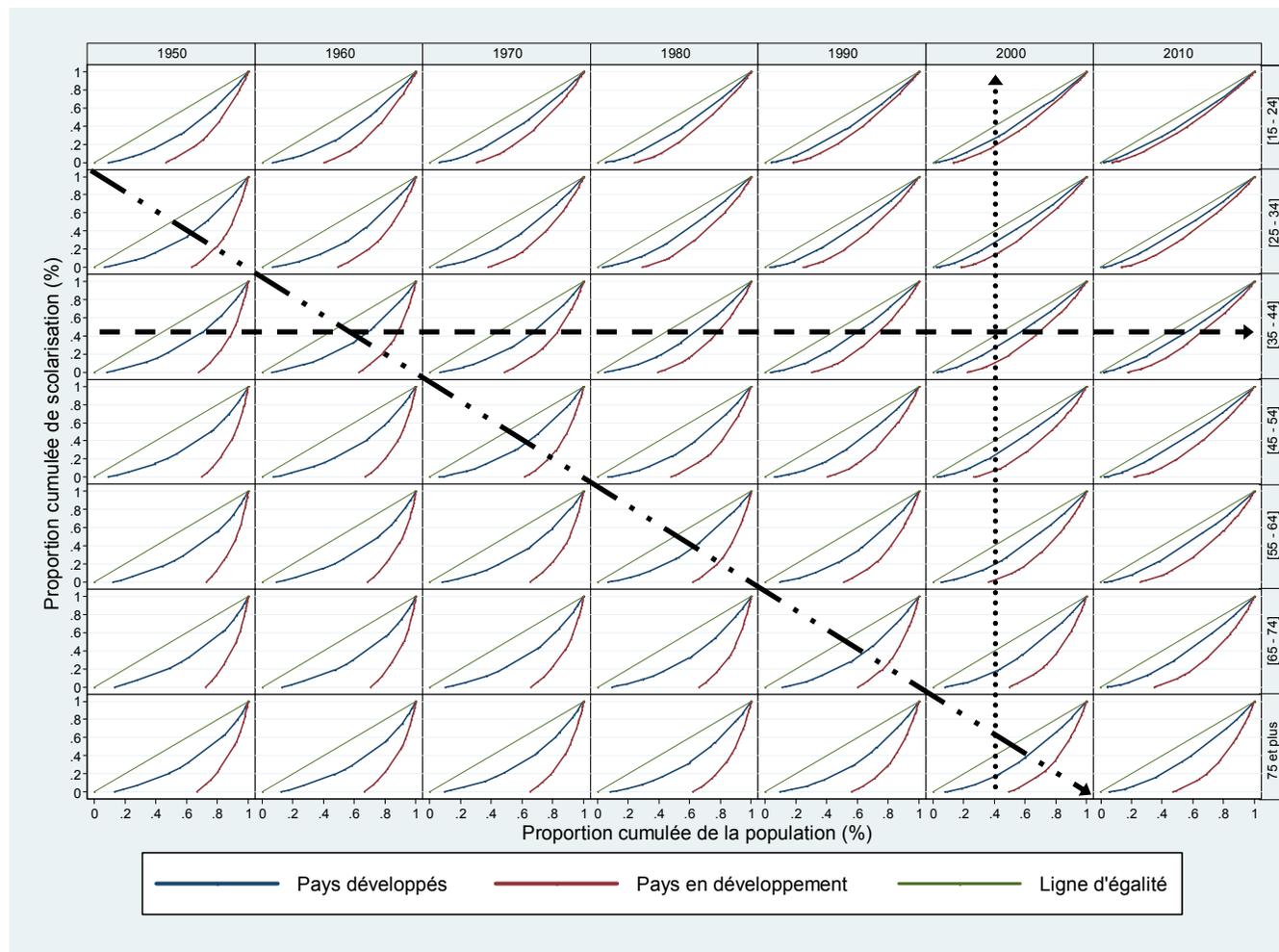


Source: Auteur (Base mondiale des inégalités éducatives)

En conséquence, la proportion de la population âgée de 15 et plus n'ayant aucune scolarisation est passée de 61.1% en 1950 à 17.4% en 2010, celle ayant atteint le niveau primaire a légèrement reculé, de 29.9% en 1950 à 26.9% en 2010, celle ayant atteint le secondaire a fortement augmenté passant de 8.1% en 1950 à 45.2% en 2010 et celle ayant atteint le supérieur est passée de 0.9% en 1950 à 10.5% en 2010. Ce qui montre que les efforts ont été plus précisément consentis dans la réduction du nombre d'analphabètes et l'augmentation de la part de la population ayant atteint le niveau secondaire. De plus, il faut signaler que l'évolution des courbes de Lorenz de l'éducation entre 1950 et 2010 confirme une convergence relative entre les pays développés et les pays en développement (graphique n° 8).

Le graphique n° 9, qui présente la courbe de Lorenz de l'éducation à la fois par niveau de développement et par groupes d'âge, confirme ce constat pour les différents groupes d'âge. La lecture horizontale du graphique n° 9 montre au cours des six dernières décennies un rapprochement entre pays développés et pays en développement au sein de chacune des tranches d'âge. Comme le montre l'exemple 1 du graphique n° 9, pour la même tranche d'âge de 35-44 ans, l'écart entre la courbe de Lorenz de l'éducation des pays développés et des pays en développement ne cesse de baisser depuis 1950. De même une lecture verticale du graphique n° 9 témoigne aussi, pour chaque année, d'un rapprochement entre les pays en développement et les pays développés en matière de réduction des inégalités éducatives. L'exemple 2 du graphique n°9 illustre, pour une coupe transversale de l'année 2000, que l'écart entre les deux courbes de Lorenz se rétrécit au fur et à mesure que l'on passe d'une tranche d'âge à une tranche inférieure.

Graphique 9 : La courbe de Lorenz de l'éducation par groupes d'âge et par niveau de développement, 1950-2010



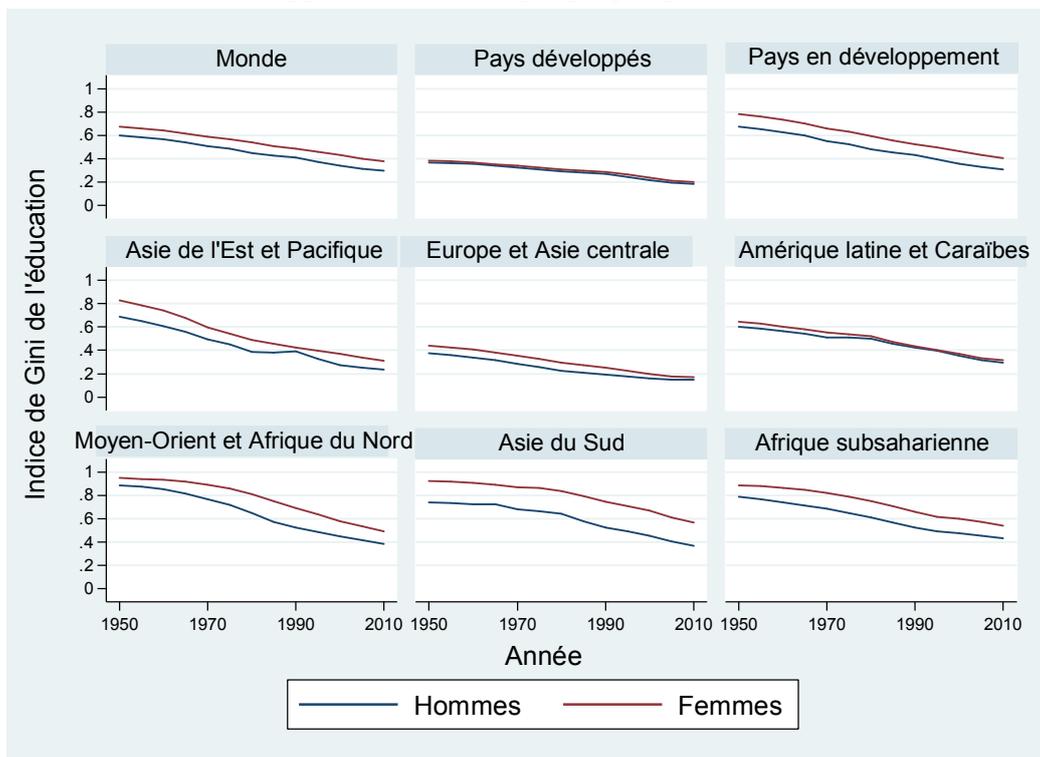
- - - → Exemple 1 : Tranche d'âge [35-44] de 1950 à 2010 : Dynamique temporelle de rapprochement entre pays développés et en développement.
-→ Exemple 2 : Année 2000, de la tranche d'âge 75 ans et plus à la tranche [15-24] : Dynamique générationnelle de rapprochement entre pays développés et en développement.
- . . → Exemple 3 : Tranche d'âge [25-34] en 1950 : Dynamique d'une cohorte.

Source: Auteur (Base mondiale des inégalités éducatives).

Par ailleurs, une lecture diagonale du graphique n°9 fait apparaître une diminution de l'écart entre pays développés et en développement pour la même cohorte (exemple 3 du graphique n° 9). L'écart reste presque fixe jusqu'à l'âge de 75 ans, au-delà duquel on assiste à un léger rapprochement dû essentiellement à deux facteurs qui se conjuguent : (i) la part élevée des illettrés chez les personnes âgées des pays en développement, (ii) un taux de mortalité relativement plus élevé chez les personnes âgées ayant de faibles niveaux d'éducation.

Le graphique n° 10 synthétise les évolutions des inégalités éducatives par genre au cours des six dernières décennies selon le niveau de développement et la zone géographique (Moyen-Orient et Afrique du Nord, Afrique subsaharienne, Asie du Sud...). En dépit des progrès réalisés en termes de réduction des inégalités éducatives chez les femmes, l'analyse globale de l'écart de genre – mesuré par la différence absolue entre les indices de Gini chez les hommes et chez les femmes – laisse augurer un rattrapage féminin différencié selon le niveau de développement et selon la région d'appartenance.

Graphique 10 : indice de Gini de l'éducation par genre, niveau de développement et zone géographique, 1950-2010



Source: Auteur (Base mondiale des inégalités éducatives).

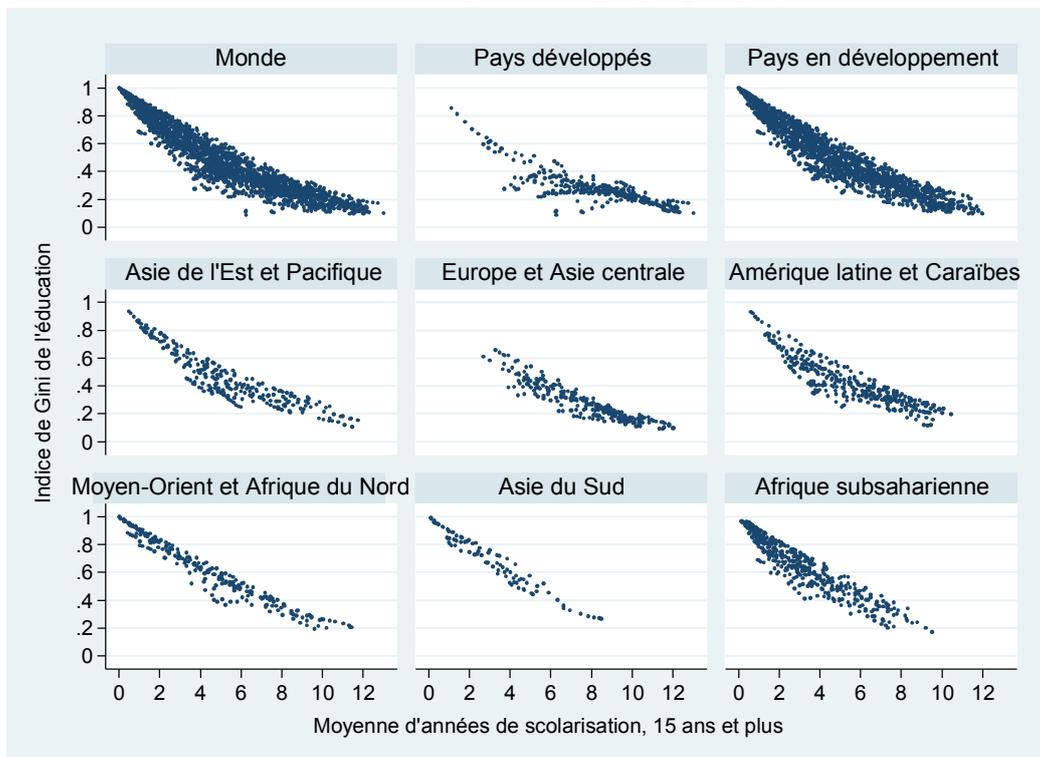
Si l'écart entre les sexes est nettement plus faible pour les pays développés, il demeure persistant pour les pays en développement (graphique n°10). Sur le plan géographique, l'écart de genre est quasi nul en Amérique latine et Caraïbes, en Europe et en Asie centrale, il est relativement moins important en l'Asie de l'Est et Pacifique. Par contre, il subsiste en Asie du Sud, en Afrique subsaharienne, et au Moyen-Orient et Afrique du Nord (20%, 11% et 11% respectivement).

Néanmoins ces régions ont réalisé des progrès non négligeables dans la réduction des inégalités éducatives aussi bien pour les hommes que pour les

femmes. Par exemple, les indices de Gini de l'éducation des femmes et des hommes de la zone MENA se situent en 1950 à des niveaux très élevés (0.95 pour les femmes et 0.89 pour les hommes) pour atteindre en 2010 des niveaux beaucoup plus bas (0.49 pour les femmes et 0.38 pour les hommes).

Il faut souligner que si l'on décline les indices de Gini de l'éducation des hommes et des femmes, à la fois par tranches d'âge, par niveau de développement et par zone géographique comme présenté dans l'annexe n°1, on constate, de manière générale, que l'écart de genre est devenu beaucoup moins marqué, notamment pour les jeunes. Pour les pays de l'Asie du Sud et de l'Afrique subsaharienne, cet écart s'est rétréci pour les jeunes de 15 à 19 ans tout au long de la période pour atteindre 11% et 5% respectivement en 2010. Pour le Moyen-Orient et Afrique du Nord, l'écart s'est estompé en 2010. Pour l'Amérique latine et Caraïbes, l'Europe, l'Asie centrale, l'Asie de l'Est et le Pacifique, il s'est même inversé au profit des jeunes filles.

Graphique 11 : lien entre l'indice de Gini de l'éducation et la moyenne d'année de scolarisation par zone géographique, 1950-2010



Source: Auteur (Base mondiale des inégalités éducatives).

Le graphique n°11 présente le lien entre l'indice de Gini de l'éducation et la moyenne d'années de scolarisation pour la population âgée de 15 ans et plus durant la période 1950-2010. Il met en évidence une corrélation négative entre ces deux variables, et ce indépendamment du niveau de développement et de la zone géographique²⁹. Ce résultat correspond à ceux des travaux empiriques traitant du lien entre l'indice de Gini de l'éducation et la moyenne d'années de scolarisation pour la population âgée de 15 ans et plus (Thomas et al., 2001). De manière

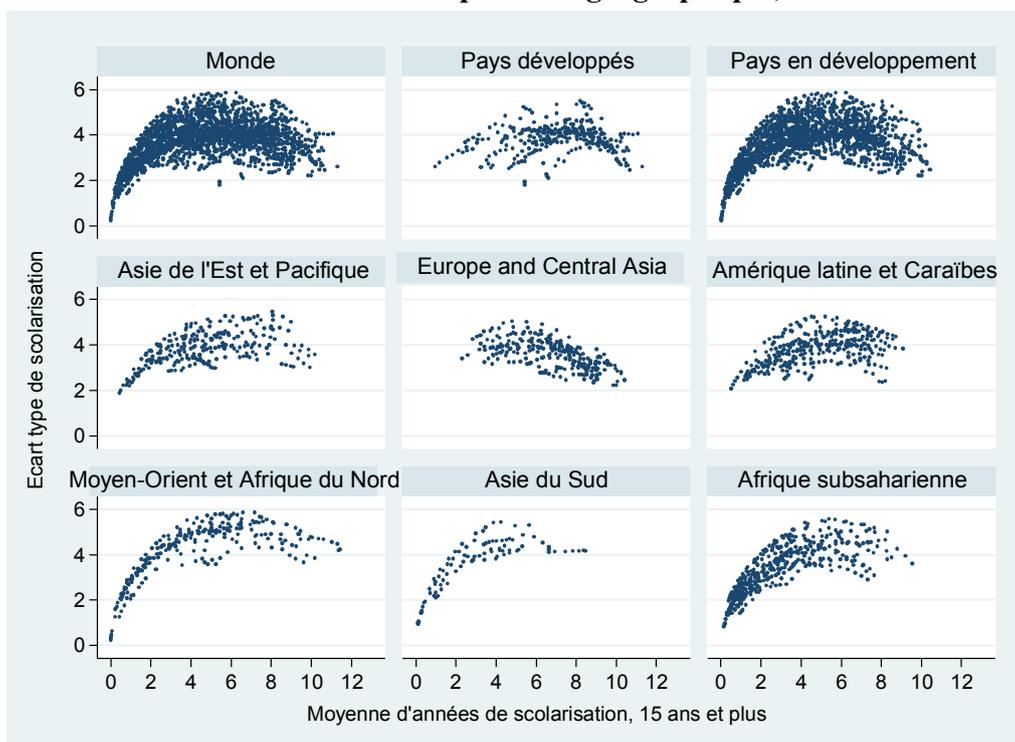
²⁹ La même tendance à la baisse entre ces deux variables est aussi observée pour l'ensemble des tranches d'âge [15-19] ans, [20-24] ans, ..., 75 ans et plus.

générale, les pays ayant un nombre d'années moyen de scolarisation suffisamment élevé pour la population appartenant à une tranche d'âge donnée sont susceptibles de connaître de faibles niveaux d'inégalités éducatives pour cette même tranche d'âge. Néanmoins, il faut souligner que cette tendance baissière globale cache des spécificités propres à certains pays. Ainsi à titre d'exemple, comme le montre l'annexe n° 2.1, pour la Belgique, la France ou encore les Pays-Bas, le lien suit une courbe en forme de cloche ("U" inversé) ↷. Tandis que pour les pays comme le Lesotho, le Malawi et la Zambie, le lien correspond plutôt à la forme en "N" inversé ↵.

Le graphique n°12 présente la relation entre l'écart-type de scolarisation et la moyenne d'années de scolarisation pour la population âgée de 15 ans et plus durant la période 1950-2010. Ce lien suit une courbe en "U" inversé à la fois pour les pays développés et en développement et pour l'ensemble des régions du monde. Ainsi, au cours du développement de la scolarisation, l'inégalité absolue d'éducation mesurée par l'écart-type de scolarisation augmente jusqu'à atteindre un certain pic, au-delà duquel elle commence à diminuer. Comme dans les travaux empiriques précédents sur cette question (Londoño, 1990; Ram, 1990 et Thomas et al., 2001), ce point de retournement est d'environ sept ans.

Il faut aussi noter que cette courbe ne suit pas toujours une forme en "U" inversé. En effet, comme le montre l'annexe n° 2.2, deux types de pays sont à distinguer : (i) certains pays, comme le Burundi, le Mali ou encore le Niger, qui ont une faible dotation éducative, n'ont pas encore atteint ce point de retournement, (ii) pour d'autres pays tels que les États-Unis, l'Arménie ou la Croatie, le lien correspond plutôt à une tendance à la baisse dès le début des années 1950.

Graphique 12 : Lien entre l'écart type de scolarisation et la moyenne d'années de scolarisation par zone géographique, 1950-2010



Source: Auteur (Base mondiale des inégalités éducatives).

1.6. Conclusion

Dans ce premier chapitre, nous avons présenté un cadre méthodologique pour la construction des indices d'inégalités éducatives. Nous avons montré l'importance de l'estimation de la variable d'éducation, de l'entité d'observation, et de la population d'intérêt sur laquelle l'inégalité est calculée. Nous avons aussi mis en évidence que le comportement vis-à-vis des changements dans la distribution de l'éducation varie d'un indice d'inégalité à l'autre. Ainsi, si le coefficient de variation et l'écart-type du logarithme sont sensibles respectivement aux extrémités supérieure et inférieure de la distribution, l'indice de Gini quant à lui est peu sensible aux deux queues de la répartition de la variable d'éducation.

Partant de l'analyse des différentes bases de données sur les inégalités éducatives à l'échelle mondiale, nous avons proposé une nouvelle base qui couvre plusieurs indices d'inégalité pour 146 pays entre 1950 et 2010. Celle-ci corrige la pondération non adéquate qui a été utilisée par Barro et Lee (2013) pour agréger les durées cumulées de scolarisation par rapport à la tranche 15 ans et plus. De plus, elle prend en considération la variation au cours du temps de ces durées au sein des différents groupes d'âge. Nous avons non seulement amélioré la méthodologie du calcul de l'inégalité d'éducation mais aussi apporté de nouvelles estimations pour des tranches d'âge qui n'étaient pas disponibles jusqu'il y a peu. Notre base de données nous a ainsi permis de présenter de nouveaux résultats par groupes d'âge de 5 ans. A titre d'exemple, à travers une comparaison des courbes de Lorenz par groupes d'âge, nous avons montré que l'écart d'inégalité d'éducation entre les pays développés et les pays en développement devient relativement moins important pour les plus jeunes.

Par ailleurs, nous avons montré que les indices d'inégalité d'éducation sont fortement liés au taux d'analphabétisme et que la baisse de l'indice de Gini de l'éducation est expliquée, plus ou moins, par la diminution du taux d'analphabétisme. Nous avons mis en évidence que cet indice devient de moins en moins dépendant du taux d'analphabétisme si ce dernier se situe à des niveaux relativement faibles.

Après avoir discuté la méthodologie suggérée par Berthélemy (2006), nous avons montré que la correction de l'indice d'inégalité d'éducation qu'il a proposé ne parvient pas à résoudre le problème de corrélation avec le niveau moyen d'éducation. Après avoir généralisé la formule proposée par cet auteur, nous avons prouvé que l'on ne peut dissocier l'indice de Gini de l'éducation du niveau moyen. Nous avons aussi mis en évidence que cet indice dépend non seulement du niveau moyen d'éducation mais aussi de l'ensemble des durées cumulées de scolarisation. A travers l'encadrement de cet indice, nous avons aussi déduit graphiquement le domaine de variation possible de ces deux variables en fonction des durées cumulées d'enseignement.

En examinant un certain nombre d'aspects d'ordre méthodologique, qu'ils soient attachés aux problèmes de pondérations non appropriées, ou qu'ils soient liés à certaines hypothèses réductrices sur les niveaux d'éducation retenus ou sur les durées des cycles d'enseignement, ce chapitre prétend apporter des éclairages sur la mesure des inégalités d'éducation. Il nous a aussi permis de disposer d'une mesure alternative aussi bien du stock moyen de l'éducation que de l'inégalité

monétaire. A partir de là, nous pouvons nous demander dans quelles conditions la diminution de l'inégalité dans l'éducation peut favoriser la croissance.

Le chapitre suivant propose d'explorer *la non linéarité* de la relation entre les inégalités d'éducation et le développement économique. Compte tenu des aspects méthodologiques présentés dans le premier chapitre, nous avons ainsi pris le choix de ne pas intégrer simultanément la moyenne d'années de scolarisation avec l'indice d'inégalité d'éducation dans l'explication du développement. En plus des variables de contrôle, nous nous contentons de retenir uniquement l'inégalité dans l'éducation tout en diversifiant sa mesure à partir de notre propre base mondiale.

Chapitre II

Inégalités éducatives et développement : l'existence d'une relation non linéaire

2.1. Introduction

2.2. Les enseignements de la revue de la littérature

2.3. La spécification non linéaire du modèle

**2.4. L'impact des inégalités éducatives sur le développement
économique : une vérification empirique de la non-linéarité**

2.5. Conclusion

Chapitre II

Inégalités éducatives et développement : l'existence d'une relation non linéaire

2.1. Introduction

L'analyse de la relation entre inégalités éducatives et développement s'appuie sur deux courants de recherche, le premier concerne le lien entre éducation et croissance et le deuxième s'intéresse plutôt au lien entre inégalité et développement.

La littérature empirique autour du premier courant est abondante et les conclusions sont souvent divergentes. En particulier, la relation éducation/croissance n'est pas toujours vérifiée dans le cas des pays en développement (Delgado et al., 2013). Pour beaucoup de ces pays, l'accès au système éducatif demeure encore limité et la recherche d'une scolarisation massive a lieu souvent aux dépens de la qualité de l'éducation reçue. Le niveau d'éducation de la population active et les conditions du développement économique peuvent être insuffisants pour que les effets bénéfiques de l'éducation sur le processus de croissance soient perceptibles.

La relation entre inégalités et développement a fait l'objet de nombreuses études notamment depuis l'hypothèse de la courbe en cloche de Kuznets (1955). Là encore, les résultats empiriques, confirmant si l'inégalité stimule ou freine la croissance, sont mitigés.

Ce manque de consensus concernant ces deux relations peut s'expliquer de différentes manières : par des insuffisances d'ordre méthodologique liées particulièrement aux mesures adoptées et/ou par des problèmes de spécification des modèles empiriques, ou encore par le choix des échantillons (mêlant des pays à niveaux de développement différents) et les périodes d'analyse considérées.

Concernant les problèmes de mesure, les indices de scolarisation ou la moyenne d'années de scolarisation sont souvent utilisés. Or, le succès de l'expansion quantitative de l'éducation dans un certain nombre de pays en développement, mesurée par les indices de scolarisation, ne doit pas occulter le degré de répartition des niveaux d'instruction entre les individus. Des écarts d'instruction peuvent exister jusqu'à compromettre la réalisation des objectifs liés à l'égalité des chances.

Le taux de scolarisation est un indicateur qui renseigne globalement sur le degré d'accès à l'éducation mais sous-estime les niveaux cumulés de scolarité, assimilés dans de nombreux travaux au stock de capital humain. Le taux d'alphabétisation des jeunes, quant à lui, ne saisit que les premiers stades de l'accumulation du capital humain et néglige, de ce fait, les connaissances et les compétences acquises au-delà de l'éducation de base. Le niveau de scolarité maximum atteint est fréquemment utilisé dans la littérature empirique pour mesurer le niveau du capital humain mais ne permet pas de caractériser la distribution de l'éducation au sein d'une population.

La relation inégalité/développement a souvent été appréhendée au travers de sa seule dimension monétaire. Les inégalités d'accès aux opportunités, particulièrement en matière d'éducation, sont considérées comme un obstacle essentiel à la justice sociale et entravent de ce fait le processus de développement (Müller, 2014). Dans les pays en développement, les inégalités éducatives sont d'autant plus importantes qu'elles sont source d'autres formes d'inégalité puisqu'elles conditionnent la distribution des revenus, la capacité d'adaptation aux changements culturels et technologiques, le niveau de bien-être, voire de cohésion sociale (Green, 2011).

Sur le plan de la spécification du modèle empirique, la relation inégalité éducative/développement a été le plus souvent testée à travers des modèles paramétriques classiques où le lien est supposé linéaire. Lorsque la relation est supposée non linéaire, elle est souvent vérifiée par des modèles incluant une forme quadratique et/ou d'ordre plus élevé (Yu et al., 2015; Güngör, 2010) ou une variable d'interaction entre l'indice d'inégalité et la moyenne d'années de scolarisation (Gille, 2015 ; Sauer et Zagler, 2014). Ce type d'approche présente l'inconvénient de restreindre la non-linéarité à une forme fonctionnelle spécifique (quadratique ou cubique notamment dans la plupart des travaux).

En utilisant un cadre méthodologique plus souple qui n'impose aucune restriction à la forme fonctionnelle (Henderson et al., 2008; Zhou et Li, 2011), nous réexaminons, dans ce chapitre, la non-linéarité de la relation moyennant des modèles non-paramétriques et semi-paramétriques combinés à l'utilisation des données de panel. Nous exposons dans une première section de ce chapitre la littérature théorique et empirique de la relation. Dans une seconde section nous présentons les données utilisées et la spécification du modèle. Nous analysons dans une troisième section les résultats obtenus faisant apparaître l'existence de relations non linéaires. Nous discutons, dans une quatrième section, à travers deux types de décomposition de l'inégalité éducative, des composantes qui contribuent le plus au développement économique.

2.2. Les enseignements de la revue de la littérature

Nous discutons dans cette section les différentes tentatives d'incorporation des inégalités éducatives dans les modèles expliquant la croissance économique. Ainsi nous présentons dans un premier temps la littérature théorique qui met en évidence les canaux à travers lesquels la distribution de l'éducation peut impacter la performance économique, puis dans un deuxième temps nous discutons les travaux empiriques qui ont analysé cette relation.

On suppose que la fonction de production de chaque individu « i » appartenant au pays « c » à l'année « t » s'écrit comme suit $y_{c,i,t} = F(k_{c,i,t}, h_{c,i,t})$ où la fonction F est croissante pour $k_{c,i,t}$ et $h_{c,i,t}$. Autrement dit, l'amélioration du capital aussi bien physique $k_{c,i,t}$ qu'humain $h_{c,i,t}$ augmente la productivité individuelle $y_{c,i,t}$. On note respectivement $\overline{k_{c,t}}$ et $\overline{h_{c,t}}$ la moyenne du capital physique et humain du pays « c » à l'année « t » telle que:

$$\overline{k_{c,t}} = \frac{1}{N} \sum_{i \in c} k_{c,i,t} \text{ et } \overline{h_{c,t}} = \frac{1}{N} \sum_{i \in c} h_{c,i,t}$$

On note aussi $K_{c,t}$ et $H_{c,t}$ le capital physique et humain du pays « c » à l'année « t » et $Y_{c,t}$ la production nationale de ce pays à la même année. Cette dernière est représentée par l'agrégation des fonctions de production individuelles selon la formule suivante :

$$Y_{c,t} = \sum_{i \in c} y_{c,i,t} = \sum_{i \in c} F(k_{c,i,t}, h_{c,i,t}).$$

Etant donné que le capital physique est transférable entre les individus, on peut définir le capital physique national comme étant la somme des capitaux physiques individuels ($K_{c,t} = \sum_{i \in c} k_{c,i,t}$). Néanmoins, ceci ne peut être valable pour le capital humain. En effet, ce dernier ne se prête pas à un parfait échange et pose un problème d'agrégation. Une formulation plus appropriée du capital humain agrégé serait différente de la somme des capitaux humains individuels: ($H_{c,t} \neq \sum_{i \in c} h_{c,i,t}$). Ainsi, on peut définir la production nationale du pays « c » à l'année « t » comme étant une fonction du capital physique national et des capitaux humains individuels :

$$Y_{c,t} = F(K_{c,t}, h_{c,1,t}, h_{c,2,t}, \dots, h_{c,N,t})$$

Nous illustrons ce problème d'agrégation à travers un exemple sur six pays fictifs A, B, C, D, E et F. Le tableau n°10 présente la répartition des individus de chacun de ces pays selon le niveau d'éducation maximum atteint. Les durées des cycles d'enseignement « y_i » de ces derniers sont supposées les mêmes et sont toutes présumées des multiples de quatre par ordre croissant (4 années pour le primaire, 8 années pour le collège, 12 années pour le lycée et 16 années pour le supérieur).

Tableau 10 : Exemple illustratif du problème d'agrégation

	y_i	Pays A	Pays B	Pays C	Pays D	Pays E	Pays F
Aucune scolarisation	0	0	25	50	55	0	10
Primaire	4	50	0	0	0	75	70
Collège	8	50	75	0	15	0	0
Lycée	12	0	0	50	0	25	0
Supérieur	16	0	0	0	30	0	20
Nombre total d'individus	-	100	100	100	100	100	100
Moyenne d'années de scolarisation	-	6,0	6,0	6,0	6,0	6,0	6,0
Indice de Gini de l'éducation	-	0,17	0,25	0,50	0,61	0,25	0,38

Source : Auteur.

Nous constatons ainsi que la somme des capitaux humains individuels est identique pour l'ensemble de ces pays. La somme des capitaux humains individuels pour l'ensemble des six pays est égale à 600 (par exemple pour le pays A c'est $50 \cdot 4 + 50 \cdot 8 = 600$ et pour le pays D c'est $55 \cdot 0 + 15 \cdot 8 + 30 \cdot 16 = 600$). Ce qui implique aussi que la moyenne d'années de scolarisation est aussi identique et égale à 6 pour l'ensemble des six pays.

Pourtant, comme le montre le tableau n°10, le même niveau de capital humain cache des niveaux d'inégalité d'éducation très dissemblables d'un pays à l'autre. En effet, les pays A et D présentent respectivement la distribution d'éducation la moins et la plus inégalitaire parmi les six pays considérés. L'indice de Gini de l'éducation est de l'ordre de 0,17 pour le pays A et de 0,61 pour le pays D. De ce fait, même si ces deux pays possèdent la même dotation éducative, la productivité marginale du capital humain agrégé du pays D est susceptible d'être

moins importante par rapport à celle du Pays A en raison de la différence dans la distribution de l'éducation au sein de ces deux pays.

Plus encore, l'utilisation d'un seul indice d'inégalité peut aussi masquer l'hétérogénéité de la composition des qualifications au sein d'une économie donnée. Comme le montre le tableau n°10, si l'on regarde par exemple les pays B et E, on constate qu'ils ont le même niveau moyen d'éducation et le même indice de Gini de l'éducation. Néanmoins, la distribution de l'éducation dans ces deux pays n'est pas la même³⁰, celle-ci fera probablement jaillir des disparités dans leurs productivités marginales. C'est pour cette raison qu'il est important de diversifier les mesures sur la distribution du capital humain pour plus de fiabilité.

Le passage du niveau microéconomique au niveau macroéconomique ne peut être systématiquement déduit au moyen d'une agrégation reposant sur la somme des capitaux humains individuels. En effet, les individus interagissent entre eux, engendrent des synergies et exercent des effets de complémentarité non négligeables dans la création de la richesse.

En se basant sur le développement de Taylor d'ordre 2 de la fonction de production et en supposant de plus que celle-ci est une fonction de type Cobb-Douglas, Lopez et al. (1998), Park (2006) et Földvári et Van Leeuwen (2011) mettent en avant deux principaux constats : (i) la distribution de l'éducation est une variable importante qui est omise dans les précédents modèles théoriques, (ii) la bonne spécification de la relation entre inégalités dans l'éducation et développement économique se révèle être non linéaire. Néanmoins, contrairement à ce qu'ils ont pu démontrer dans le modèle théorique, ces auteurs se sont appuyés sur une simple modélisation linéaire. *"The relationship ... is obviously nonlinear... For convenience, we approximate the relationship with a linear regression"* (Földvári et Van Leeuwen, 2011). *"A linearized empirical model ... is adopted to investigate whether average index and dispersion index of human capital have any significant growth implication"* (Park, 2006).

Nous discutons dans ce qui suit les différents mécanismes à travers lesquels les inégalités éducatives agissent sur la croissance et qui justifient l'importance de l'intégration de la distribution du capital humain dans les modèles expliquant la performance économique. En effet, la littérature théorique met en évidence quatre principaux mécanismes. Le premier d'entre eux concerne le modèle proposé par Galor et Zeira (1993) sur les contraintes d'investissement dans le capital humain. Ces dernières sont liées principalement aux imperfections des marchés du crédit. Les auteurs supposent qu'ils sont imparfaits, dans le sens où les asymétries d'information amènent les institutions financières à un rationnement du crédit. Ce dernier se fait généralement en faveur des individus potentiellement capables de rembourser leur emprunt. En effet, si l'on suppose que l'éducation est en majorité financée directement ou indirectement par le secteur privé, le manque de fonds propres des individus à revenu faible les empêchera d'améliorer leur niveau de qualification. Ne pouvant emprunter, ces individus, se trouvent contraints de ne pas

³⁰ A titre d'exemple la différence entre le taux d'analphabétisme des pays B et E est de l'ordre de 25% alors que les deux pays présentent la même moyenne d'années de scolarisation et le même indice de Gini de l'éducation.

pouvoir bénéficier d'un développement de leur capital humain même s'ils sont convaincus de la rentabilité d'un tel investissement.

Ainsi, puisque les marchés de crédit sont imparfaits, le financement de l'accumulation du capital humain est prédéterminé par la richesse initiale des individus. Autrement dit, une répartition plus inégalitaire de la richesse entraîne une baisse des taux d'investissement en capital humain. Cette dernière concerne aussi bien le court terme que le moyen et le long terme³¹. Galor et Moav (2004) ajoutent que l'effet négatif des inégalités sur la croissance pour les pays développés devrait être de moins en moins perceptible. Dans ces pays, les conditions rattachées à l'octroi d'un crédit d'investissement dans le capital humain sont moins contraignantes. En revanche, pour les pays en développement qui se caractérisent par des niveaux faibles de capital humain et des frais d'éducation assez élevés, l'effet négatif est beaucoup plus présent.

Contrairement au mécanisme des imperfections des marchés du crédit, qui sous-tend un financement majoritairement privé de l'éducation, les modèles de l'économie politique considèrent aussi bien l'enseignement éducatif public que privé. Les premiers travaux en ce sens (Alesina et Rodrik, 1994; Persson et Tabellini, 1994) montrent que les sociétés caractérisées par une répartition inégale de la richesse votent en faveur de politiques redistributives inefficaces qui affectent négativement la croissance économique. D'autres travaux comme celui de Benabou (2000) se réfèrent au comportement électoral des individus dans le choix du régime éducatif. Ces modèles considèrent plutôt que l'effet des politiques redistributives sur la performance économique serait dépendant de l'écart entre la qualité de l'enseignement privé et public.

En effet, si l'inégalité d'éducation est suffisamment prononcée de telle sorte que la proportion des individus ayant un capital humain en dessous de la moyenne soit trop élevée, la médiane de la distribution serait dans ce cas inférieure à la moyenne. Ainsi, le choix de l'électeur médian se traduira-t-il en faveur d'un enseignement éducatif public. Ce dernier, financé par les impôts, permet au moins aux individus pauvres d'accumuler un certain niveau d'éducation. En revanche, si l'inégalité d'éducation est relativement faible, la proportion des individus accumulant un niveau de capital humain élevé est plus forte. Par conséquent, l'électeur médian choisira le régime de l'enseignement privé, qui favorisera l'investissement dans la qualité de l'enseignement.

Dans ce cas, l'effet à court terme des politiques redistributives sur la croissance économique serait dépendant de la qualité de l'enseignement public. Par la suite, les inégalités d'éducation, établies à court terme, sont transmises aux générations suivantes, à travers la relation « capital humain/revenu », et sont ainsi largement reproduites à moyen et à long terme.

Un troisième mécanisme tente d'expliquer la relation à travers l'instabilité socio-politique. Dans un contexte caractérisé par de fortes inégalités et un environnement social tendu, les couches vulnérables expriment leur mécontentement social par des manifestations et des revendications d'égalité. Ainsi le climat socio-politique incertain affaiblit la capacité du gouvernement à réagir

³¹ La distribution des revenus affecte également la répartition de la richesse des générations ultérieures.

face à cette instabilité ce qui impacte négativement la performance économique du pays.

Certains travaux ont avancé un quatrième mécanisme qui est celui du canal démographique. Les écarts individuels par rapport aux variables démographiques, comme la fertilité ou l'espérance de vie, peuvent influencer l'accumulation du capital humain. De la Croix et Doepke (2003) analysent l'effet des inégalités sur la croissance à travers la fécondité. Dans ce modèle, les parents pauvres ayant un faible niveau d'éducation décident d'avoir plus d'enfants. Ces enfants, étant nombreux issus d'une famille pauvre, ne peuvent bénéficier que de faibles niveaux d'éducation. En revanche, les enfants de parents relativement aisés sont moins nombreux et ont des niveaux d'éducation relativement élevés. On s'attend donc à ce que ce différentiel de fécondité entre parents plus et moins instruits se répercute sur la proportion d'individus moins instruits dans les générations futures. Ceci implique à son tour un faible taux d'investissement en capital humain.

Si le lien entre inégalités éducatives et développement emprunte son fondement théorique à des courants assez anciens³², la littérature empirique de ce lien au niveau transnational est toutefois relativement récente. Ceci est notamment dû, comme l'ont souligné Castelló et Doménech (2002), au manque de bases de données sur les inégalités d'éducation comparables au niveau international. *“Due to the lack of available data on human capital inequality, little attention has been devoted to the influence of human capital distribution on economic growth in empirical studies”* (Castelló et Doménech, 2002). De même Johnes (2006) note aussi à cet égard *“So what is, then, the relationship between educational inequalities and growth? Until the last few years, this was not a question that could be answered”*.

Nous présentons dans ce qui suit les travaux empiriques qui ont abordé la relation au niveau transnational³³. En utilisant des données en coupe transversale sur quarante-trois pays en développement, Birdsall et Londoño (1997) sont les premiers à avoir étudié la relation inégalités éducatives et croissance économique. Ils ont analysé l'impact de l'inégalité d'éducation, mesurée par l'écart-type de scolarisation, sur la croissance du revenu par habitant et la réduction de la pauvreté entre 1973 et 1982. Les auteurs constatent que l'effet est négatif et significatif et qu'il est d'autant plus marqué pour les pauvres.

Contrairement à Birdsall et Londoño (1997) qui ont testé la relation sur des données en coupe transversale moyennant une mesure absolue de l'inégalité, Lopez et al. (1998) ont utilisé des données de panel et des mesures relatives de l'inégalité. Les auteurs arrivent à la même conclusion. En effet, en utilisant le coefficient de variation et l'écart-type du logarithme de scolarisation sur douze pays en développement entre 1970 et 1994, ils ont identifié une relation négative entre le niveau de l'inégalité d'éducation et le PIB par habitant. Les auteurs défendent l'idée que le faible niveau de croissance enregistré dans les pays en développement malgré les efforts consentis dans la massification de l'éducation est principalement dû à la répartition asymétrique de l'éducation. Lopez et al. (1998) ne contestent pas

³² Le premier concerne la relation inégalité/développement qui remonte à la fameuse courbe en cloche de Kuznets (1955) tandis que le second s'est développé depuis la théorie du capital humain de Becker (1964).

³³ La littérature empirique de la relation discutée dans ce chapitre concerne uniquement les approches transnationales. La littérature au niveau infranational fera l'objet d'une présentation au chapitre III.

l'importance d'une dotation moyenne élevée de capital humain, mais suggèrent que cet aspect ne saurait à lui seul garantir une croissance soutenue, encore faut-il que ce capital humain soit équitablement réparti dans la population.

D'autre part, Castelló et Doménech (2002) et Bowman (2007) ont aussi testé la relation entre les inégalités dans l'éducation mesurée par l'indice de Gini de l'éducation et la croissance économique pour deux échantillons en coupe transversale de 108 et 57 pays. Contrairement à Lopez et al. (1998) qui se réfèrent dans leur explication à la relation entre éducation et croissance, ces auteurs renvoient au lien entre inégalité et croissance. En effet, ils défendent la thèse selon laquelle la distribution du capital humain, en comparaison à celle du revenu, apporte des estimations plus robustes quant à l'effet de l'inégalité sur la croissance.

Néanmoins, les travaux de Birdsall et Londoño (1997), Castelló et Doménech (2002) et Bowman (2007) présentent l'inconvénient d'utiliser des données en coupe transversale. Ces dernières sont souvent critiquées dans les modèles de croissance. L'échantillon de Lopez et al. (1998) malgré l'utilisation des données de panel, concerne uniquement douze pays en développement, ce qui demeure assez réduit. L'utilisation des données de panel pour un ensemble plus large de pays permettent une meilleure robustesse.

L'un des avantages essentiel de l'utilisation des données de panel par rapport aux autres types de données (séries temporelles ou coupes transversales) réside dans le fait que, le nombre d'observations étant plus grand, le risque de colinéarité est plus faible. Contrairement aux études en coupe transversale, cette approche permet d'étudier plus précisément la dynamique de long terme de la relation. Par ailleurs, il est possible par ce biais de contrôler l'hétérogénéité individuelle ou temporelle non observée qui pourrait être corrélée avec les variables explicatives et biaiser les estimations.

Ainsi, plus récemment, la construction de bases de données transnationales mesurant les inégalités éducatives a permis l'émergence de travaux employant de larges panels de pays avec des périodes suffisamment longues. Les travaux de Thomas et al. (2001) et Park (2006) sont les premiers en ce sens. Ils ont respectivement concerné 72 pays entre 1975 et 1990 et 94 pays entre 1960 et 1995. Toutefois, ces derniers aboutissent à des conclusions différentes. En effet, le premier atteste d'une liaison négative entre l'indice de Gini de l'éducation et l'accroissement absolu du PIB par habitant³⁴ tandis que le second confirme une relation positive entre l'écart-type de scolarisation et la croissance économique.

Par ailleurs, ces divergences peuvent en partie être expliquées par les mesures mobilisées dans ces travaux. En effet, Thomas et al. (2001) utilisent une mesure relative pour l'inégalité d'éducation et une mesure absolue pour le développement économique (accroissement absolu du PIB). À l'inverse, Park (2006) emploie quant-à-lui une mesure absolue de l'inégalité d'éducation et une autre relative pour le développement économique.

³⁴ Thomas et al. (2001) suggèrent qu'une meilleure mesure du développement économique serait l'accroissement absolu du PIB par habitant en PPA.

D'autre part, malgré le fait que l'hypothèse de la non-linéarité n'ait pas été explicitement testée³⁵ par Thomas et al. (2001) et Park (2006), elle a cependant été déduite indirectement par ces derniers en s'appuyant sur l'agrégation des fonctions de production individuelles. En effet, Thomas et al. (2001) précisent qu'une fois la moyenne d'années de scolarisation et l'indice de Gini de l'éducation sont simultanément introduits, les estimations empiriques dévoilent que l'indice d'inégalité devient non significatif. Une explication³⁶ avancée par les auteurs réside dans la non-linéarité de la relation. Cette dernière est principalement due, selon eux, au problème d'agrégation³⁷ *“First, the theory tells us that there is an aggregation problem when marginal product of education for each individual is not equalized, and the function form for aggregation might be nonlinear. But empirically we have only estimated a linear function”* (Thomas et al., 2001).

En introduisant à la fois la moyenne et l'écart-type de scolarisation dans le modèle, les résultats empiriques de Park (2006) soutiennent l'idée que, étant donné le même niveau moyen de scolarisation, la composition du capital humain agrégé aurait de plus un effet supplémentaire sur la croissance. Ceci atteste, selon Park (2006), de la convexité du rendement de l'éducation et par conséquent de la conformité avec la non-linéarité évoquée dans la littérature théorique : *“This evidence is consistent with the theoretical discussion regarding the existence of non-linearity in human capital production technology”*.

Par ailleurs, Földvári et Van Leeuwen (2011) ont testé l'impact de la distribution de l'éducation à la fois sur les inégalités de revenu et sur le PIB par habitant. Pour ce qui est de l'effet sur le PIB par tête qui nous intéresse, leurs estimations confirment que l'effet diffère selon les bases de données utilisées pour mesurer l'inégalité d'éducation. En se basant sur les données de Barro et Lee (2001), les auteurs confirment que, toutes choses égales par ailleurs, la répartition inégale de l'éducation impacte négativement le niveau du PIB par habitant alors que l'utilisation des données de Cohen et Soto (2007) témoignent d'un effet positif non significatif.

Si Földvári et Van Leeuwen (2011) suggèrent que l'effet est tributaire du choix des données mobilisées, Castelló (2010b) soutient, quant à lui, un effet dépendant du niveau de développement des pays en question. D'une part, pour les pays en développement (70 pays sur la période 1965-2005), l'impact négatif est confirmé même en diversifiant la mesure de l'inégalité d'éducation. De plus, pour le cas de ces pays, l'auteur met davantage en avant « le canal démographique » parmi l'ensemble des canaux par lesquels transite la relation.

D'autre part, moyennant un panel de 32 pays développés entre 1965 et 2005, Castelló (2010b) constate des résultats contrastés concernant l'effet de l'inégalité dans l'éducation sur la croissance économique. En effet, une plus grande inégalité du capital humain réduit les taux de croissance dans les pays européens tandis qu'elle impacte positivement les économies anglo-saxonnes. Plus encore, l'auteur

³⁵ Elle n'a pas été testée ni par une spécification paramétrique en incluant une variable d'interaction ou une forme quadratique et/ou d'ordre plus élevé, ni par une spécification non-paramétrique.

³⁶ La seconde explication signalée par les auteurs réside dans l'éventuelle corrélation entre la moyenne et sa distribution. Cette dernière a aussi été mentionnée par Berthélemy (2006) : *“when human capital level and human capital concentration are introduced simultaneously in growth equations, parameters cannot be robustly estimated, due to the very high correlation between these two variables”*.

³⁷ Ce problème a aussi été soulevé par Berthélemy (2006).

soupçonne même que l'effet peut être changé d'un pays développé à l'autre. *“In the sample of higher-income countries, in contrast, no clear effect of human capital inequality on growth is found. However, a closer look shows that the lack of a clear effect in wealthy economies may be due to the fact that even in the rich countries, human capital inequality affects growth differently from one country to the next”* (Castelló, 2010b).

Sauer et Zagler (2014) ont aussi testé la relation sur un panel de 134 pays entre 1950 et 2005. Ils confirment un effet différencié selon le niveau d'éducation. En introduisant un terme d'interaction entre la moyenne d'années de scolarisation et l'indice de Gini de l'éducation, les auteurs concluent que l'impact de la répartition inégale de l'éducation sur la croissance économique est non linéaire. Ce dernier est négatif pour les pays ayant une moyenne d'années de scolarisation suffisamment élevée et positif pour les pays ayant une moyenne d'années relativement faible.

Certains travaux ont étudié la relation en faisant des comparaisons entre les régions du monde. Klasen et Lamanna (2009) montrent que les inégalités d'éducation sont à l'origine d'une différence de 0.9 à 1.7 points de pourcentage du taux de croissance dans les régions de l'Asie de l'Est, l'Asie du Sud, le Moyen-Orient et l'Afrique Sub-saharienne. De même, Birdsall et Londoño (1997) ont constaté qu'entre 1973 et 1982, les pays de l'Asie de l'Est et du Pacifique ont enregistré une croissance économique cinq fois plus élevée que celle des pays de l'Amérique latine et des Caraïbes. Ceci serait principalement dû, selon les auteurs, à la différence dans la répartition initiale du capital humain au sein des pays de ces deux régions.

Baliamoune-Lutz et McGillivray (2009, 2015) se sont intéressés à la relation pour les pays de la région MENA et de l'Afrique subsaharienne. En utilisant l'écart de genre pour les taux de scolarisation au primaire et au secondaire comme indicateur d'inégalité éducative, les résultats de leurs estimations indiquent que ces variables ont un effet négatif et significatif sur le logarithme du PIB par habitant, en particulier dans les pays de la région MENA. Pour repérer l'existence d'une non-linéarité de la relation, les auteurs ont introduit un coefficient d'interaction entre l'indicateur d'inégalité d'éducation et l'ouverture commerciale. Ce dernier est positif et statistiquement significatif, indiquant que l'inégalité dans l'éducation semble avoir un effet supplémentaire dans les économies relativement ouvertes.

Ayesha et Kashif (2015) ont analysé la relation pour les pays de l'Asie du Sud entre 1980 et 2010 tandis que Masakure (2013) s'est intéressé aux pays de l'Afrique subsaharienne (12 pays entre 1990 et 2010). Ces deux travaux attestent d'un effet négatif et significatif de l'indice de Gini de l'éducation sur le niveau du PIB par habitant. De plus, en utilisant les variables « fertilité » et « urbanisation », Masakure (2013) affirme aussi la présence d'un effet indirect de la relation dans ces pays. Ibourk et Amaghous (2013) ont aussi analysé le cas de la région MENA. Ils confirment qu'au sein même de cette région, l'effet n'est pas le même si l'on considère les pays à niveau de revenu moyen ou élevé (pays du Golfe) : *“...thereby confirming the heterogeneous performances of the countries of the MENA region as to the impact of educational inequality on economic growth”*.

2.3. La spécification non linéaire du modèle

Nous présentons dans cette section les données et la spécification du modèle utilisées pour analyser le lien entre inégalités éducatives et développement économique. Notre échantillon est composé de 120 pays en développement observé sur intervalles de cinq ans sur la période 1955-2010. Cet échantillon fera aussi l'objet d'une analyse par grandes zones géographiques. Les résultats concernant les 24 pays développés seront aussi présentés en guise de comparaison.

2.3.1. Les variables retenues

Le modèle estimé de manière générale est le suivant :

$$y_{it} = f(\text{IneqEdu}_{it-5}, \text{DepGov}_{it-5}, \text{Invest}_{it-5}, \text{Ouverture}_{it-5}, \text{PopGrowth}_{it-5})$$

- **La variable expliquée**

La variable dépendante « y_{it} » est le logarithme du PIB par habitant en PPA \$ constants 2005 issu de « PWT 7.0 », où i représente les différents pays en développement considérés et t les périodes de cinq ans entre 1955 et 2010.

- **La variable explicative de l'inégalité dans l'éducation**

Notre principale variable explicative, qui est l'inégalité d'éducation (IneqEdu), est issue de la base mondiale des inégalités d'éducation proposée dans le chapitre I. Les indices de Gini, de Theil, d'entropie généralisée d'ordre 0.5 et d'Atkinson d'ordre 1 concernent la population âgée de 15 ans et plus.

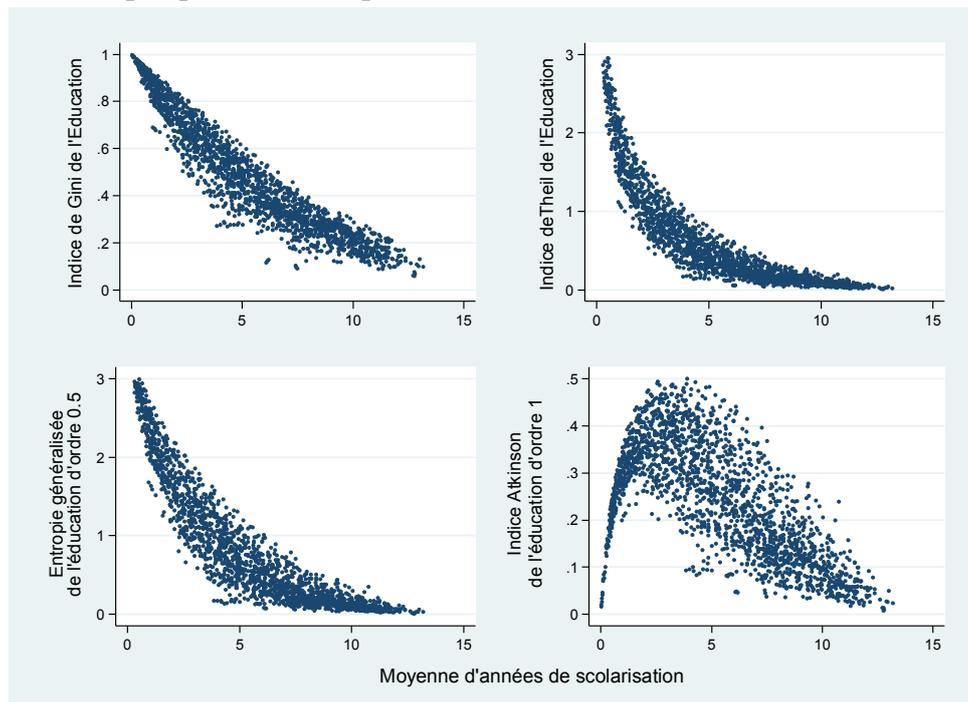
Nous avons aussi mené des décompositions (*within* et *between*) de ces quatre indices. Celles-ci ont aussi été calculées pour l'ensemble du panel et pour la même période en utilisant la méthodologie présentée dans le chapitre I. Il s'agit précisément de deux décompositions de l'inégalité d'éducation, la première concerne deux groupes (les illettrés et les non illettrés) et se décompose en une inégalité intra-éduqués et une inégalité inter-éduqués-illettrés et la seconde concerne treize groupes (13 tranches d'âge allant de [15-19] à [75 et plus]) et se décompose en une inégalité intragénérationnelle et une inégalité intergénérationnelle.

- **Le cas particulier de l'indice d'Atkinson de l'éducation**

Contrairement aux autres indices d'inégalité d'éducation retenus, l'indice d'Atkinson d'ordre 1 paraît revêtir une certaine spécificité dans la mesure où il ne montre pas une tendance à la baisse au fil du temps (graphique n° 13). C'est le seul indice à présenter une courbe en cloche entre la moyenne d'années de scolarisation et l'indice d'inégalité d'éducation. Ainsi, lorsque le niveau moyen d'éducation est très faible, l'inégalité d'éducation mesurée par l'indice d'Atkinson d'ordre 1 est plutôt faible et plus le niveau moyen d'éducation augmente, plus cet indice d'inégalité augmente jusqu'à atteindre un certain pic. Au-delà de ce pic, la tendance globale de ce lien rejoint celle des autres indices d'inégalité de l'éducation {Gini, Theil et GE(0.5)} (graphique n° 13). Il convient de noter que cette relation

particulière est principalement expliquée par le mode de calcul de cet indice. En effet, comme on l'a vu dans le chapitre I, l'indice d'Atkinson d'ordre 1 correspond au rapport entre la moyenne géométrique et la moyenne arithmétique. Étant donné que la moyenne géométrique n'accepte pas des valeurs nulles³⁸ ($y_i = 0$), la distribution de l'éducation subit au préalable une transformation avant le calcul de cet indice. Celle-ci consiste à ajouter une année de scolarisation à l'ensemble de la distribution de l'éducation pour éviter le problème des valeurs nulles ($y_i \rightarrow y_i + 1$). Cette même transformation a été adoptée par le PNUD pour évaluer la perte due à l'inégalité dans la « composante éducation » de l'indice de développement humain ajusté aux inégalités.

Graphique 13 Le cas particulier de l'indice Atkinson d'ordre 1



Source : Auteur.

- **Les variables de contrôle**

Quatre autres variables explicatives, issues de « PWT 7.0 », sont utilisées comme des variables de contrôle. On note X la matrice de ces variables où $x'_{it} = (x_{1t}, x_{2t}, x_{3t}, x_{4t})$. x_{1t} est la dépense du gouvernement liée à la consommation des administrations en pourcentage du PIB (*DepGov*). x_{2t} est le taux d'investissement mesuré par le ratio des dépenses d'investissement publiques et privées en pourcentage du PIB (*Invest*). x_{3t} est le degré d'ouverture mesurée par la somme des exportations et des importations rapportée au PIB (*Ouverture*). x_{4t} est l'accroissement de la population dans la période entre $t-5$ et t . La variable x_{4t} est calculée en adoptant la même formule utilisée dans les modèles de croissance : le logarithme népérien du taux de croissance démographique (η) plus 0.05 $\{\text{PopGrowth} = \ln(\eta+0.05)\}$.

³⁸ Contrairement à l'indice d'Atkinson d'ordre 1, l'indice de Theil même s'il n'est pas compatible avec les valeurs nulles, accepte néanmoins une limite finie lorsque les valeurs nulles sont considérées comme étant infiniment petites (y_i tend vers 0).

Etant donné que la croissance économique peut en retour affecter les inégalités éducatives, les dépenses des administrations, l'investissement, ou l'ouverture commerciale, nous avons retardé toutes les variables explicatives d'une période (de 5 années) afin de remédier aux problèmes d'interdépendance.

2.3.2. La spécification du modèle estimé

Il faut signaler que malgré le fait que certains auteurs ont mentionné le caractère non linéaire de la relation que nous étudions, aucune tentative de vérification empirique via des modèles non-paramétriques n'a été menée dans ce sens. Nous proposons ainsi dans ce chapitre d'utiliser des modèles non-paramétriques et semi-paramétriques dans le cas des données de panel pour tester la non linéarité du lien entre inégalités éducatives et développement économique.

L'intérêt de ces modèles est qu'ils permettent d'analyser le lien entre une variable dépendante et une ou plusieurs variables explicatives, sans exiger de forme prédéterminée pour la relation. Si la modélisation paramétrique s'intéresse à l'estimation d'un nombre fini de paramètres - correspondant au « nombre de variables explicatives + 1 » -, la modélisation non-paramétrique, appelée aussi « régression sans paramètres », permet d'obtenir un meilleur ajustement du nuage de points de la relation. En effet, cette modélisation présente un intérêt particulier puisqu'elle utilise des techniques d'estimation qui n'imposent aucune restriction à la forme fonctionnelle de la relation et n'exigent pas une connaissance préalable de la nature de la relation, celle-ci est dérivée directement des estimations réalisées avec moins d'hypothèses. Ainsi, dans une modélisation non-paramétrique, nous cherchons à estimer une fonction $g(z)$ appelée « fonction de lien ». Cette dernière remplace l'estimation d'un certain nombre de paramètres dans la modélisation paramétrique.

- ***Les techniques d'estimation non-paramétriques traditionnelles***

Nous donnons dans un premier temps un bref aperçu sur les premières techniques d'estimation non-paramétrique. Celles-ci concernent le lissage par opérateur à noyau appelé aussi « estimateur Kernel » dont les premiers travaux sont ceux de Nadaraya (1964) et Watson (1964) ainsi que le lissage local polynomial (loess, lowess) initié par Cleveland (1979) et Cleveland et Devlin (1988). Dans un second temps, nous présentons la technique d'estimation non-paramétrique et semi-paramétrique que nous retenons pour ce chapitre.

Le lissage par opérateur à noyau développé par Nadaraya (1964) et Watson (1964) s'inspire du lissage par moyenne mobile. Toutefois, dans ce cas la fonction de lien $g(z)$ n'est pas équipondérée. En effet, pour le cas du lissage par moyenne mobile la pondération des points z_i est la même. Celle-ci est égale à l'inverse de la fenêtre k alors que pour l'estimateur Kernel, plus les points z_i sont proches de z , plus le poids de ces derniers sera important. La fonction de lien $g(z)$ de cet estimateur se définit comme une moyenne pondérée des observations y_i où le poids est $w_i(z) = \frac{K([z_i-z]/\lambda)}{\sum_{j=1}^N K([z_j-z]/\lambda)}$ tel que λ est le paramètre de lissage. Ainsi, l'estimateur à noyau de la fonction de lien évaluée au point (z) est défini par :

$$\hat{g}(z) = \sum_{i=1}^N w_i(z) y_i$$

L'estimateur à noyau présente un avantage par rapport à l'estimateur de la moyenne mobile. En effet, ce dernier est convergent en probabilité lorsque trois conditions sont vérifiées : la taille de la fenêtre k est suffisamment grande, le rapport de celle-ci sur la taille de l'échantillon est suffisamment petit et la taille de l'échantillon est suffisamment grande. L'estimateur Kernel est en revanche convergent en probabilité dès lors que la taille de l'échantillon est suffisamment grande et non pas sous une condition supplémentaire sur le paramètre de lissage.

Une deuxième technique d'estimation concerne le lissage local polynomial (loess, lowess). Cette méthode consiste à mener une régression linéaire à chaque voisinage $V(z)$ d'un point z . Les paramètres issus de cette régression varient suivant le point de référence z . Cette méthode non paramétrique d'estimation a été initialement proposée par Cleveland (1979). Celle-ci se présente comme suit :

$$\left\{ \begin{array}{l} \hat{g}(z) = \hat{\beta}_1(z) + z\hat{\beta}_2(z) \\ \{\hat{\beta}_1(z), \hat{\beta}_2(z)\} = \underset{\{\beta_1(z), \beta_2(z)\}}{\text{MIN}} \sum_{z_i \in V(z)} [y_i - \beta_1(z) - z_i\beta_2(z)]^2 \end{array} \right.$$

Cette méthode, appelée « régression loess », consiste à identifier les paramètres $\beta_1(z)$ et $\beta_2(z)$ de la régression de manière à minimiser les moindres carrés à chaque voisinage du point (z). Ainsi ces paramètres varient selon le point de référence (z). Dans ce programme toutes les observations ont le même poids.

La technique d'estimation de Cleveland et Devlin (1988), appelée « régression lowess », généralise celle proposée par Cleveland (1979). Cette technique consiste à pondérer chaque point z_i par une fonction de poids de type Kernel. Les paramètres $\beta_1(z)$ et $\beta_2(z)$ de la régression locale sont déterminés par le programme suivant :

$$\left\{ \begin{array}{l} \hat{g}(z) = \hat{\beta}_1(z) + z\hat{\beta}_2(z) \\ \{\hat{\beta}_1(z), \hat{\beta}_2(z)\} = \underset{\{\beta_1(z), \beta_2(z)\}}{\text{MIN}} \sum_{z_i \in V(z)} K\left(\frac{|z_i - z|}{\lambda}\right) [y_i - \beta_1(z) - z_i\beta_2(z)]^2 \end{array} \right.$$

- ***Les modèles non-paramétriques et semi-paramétriques en données de panel***

Par la suite, la littérature sur l'économétrie non-paramétrique s'est développée (Rodríguez-Poo et Soberon, 2017; Lv et al., 2017) et a concerné aussi bien la régression univariée que celle multivariée. En effet, dans le cas univarié, le paramètre z de la fonction de lien $g(z)$ désigne une seule variable explicative tandis que dans le cas multivarié, z est définie comme étant un vecteur de dimension « p » où p est le nombre de variables explicatives retenues dans le modèle tel que $p > 1$. Dans ce dernier cas, on s'intéresse à un ajustement de relations multidimensionnelles où la non linéarité due aux interactions entre ces variables explicatives est aussi prise en compte.

Bien que l'introduction de l'estimation non-paramétrique date depuis les années cinquante (Rosenblatt, 1956), l'utilisation des données de panel dans ce type de modélisation est relativement très récente. Les premiers travaux en ce sens sont ceux de Li et Stengos (1996) et Li et Ullah (1998). Henderson et al. (2008) a proposé une nouvelle technique d'estimation d'un modèle non-paramétrique et semi-paramétrique sur données de panel. Celle-ci est basée sur une procédure d'estimation itérative. Zhou et Li (2011) l'ont généralisé pour le cas d'un panel non cylindré.

Dans ce chapitre, nous retenons la spécification non-paramétrique et semi-paramétrique en données de panel proposée par Zhou et Li (2011) car elle est la mieux adaptée à la nature de nos données (panel non cylindré). Puisque nous sommes amenés à estimer, dans la dernière section de ce chapitre, des surfaces non-paramétriques³⁹, nous nous plaçons dans le cas général multivarié où z désigne un vecteur de dimension « p ». La formulation pour le cas univarié est déduite en supposant « $p=1$ ». Ce cadre méthodologique plus souple permet de tester la non-linéarité de la relation en bénéficiant à la fois de l'intérêt des données de panel et de l'avantage des modèles non-paramétriques.

$$y_{it} = g(z_{it}) + u_i + \varepsilon_{it} \quad t = 1, 2, \dots, m_i; \quad i = 1, 2, \dots, n. \quad (1)$$

$$y_{it} = g(z_{it}) + x'_{it}\gamma + u_i + \varepsilon_{it} \quad t = 1, 2, \dots, m_i; \quad i = 1, 2, \dots, n. \quad (2)$$

Les spécifications (1) et (2) représentent respectivement les modèles non-paramétrique et semi-paramétrique sur données de panel à effets fixes. Nous désignons « y » la variable dépendante et « z » le vecteur de « p » variables explicatives où la fonction de lien $g(\cdot)$ qui relie le vecteur z avec la variable y est une fonction non spécifiée à estimer définit $\mathfrak{R}^p \rightarrow \mathfrak{R}$. Pour le cas du modèle semi-paramétrique, « q » d'autres variables de contrôle « x » sont considérées dont γ est un vecteur de « q » paramètres à estimer. On considère le cas d'un panel non cylindré⁴⁰ où chaque pays « i » dispose de m_i observations. Les effets individuels u_i sont considérés fixes et corrélés avec z où la forme de cette corrélation est non spécifiée. Les termes d'erreurs ε_{it} sont supposés être i.i.d, de moyenne nulle et de variance égale σ_ε^2 où $E(\varepsilon_{it} | z_{it}) = 0$. On note I_k la matrice identité de dimension k et e_k un vecteur unitaire $k \times 1$. Si on définit $\tilde{\varepsilon}_i = (\tilde{\varepsilon}_{i2}, \dots, \tilde{\varepsilon}_{im_i})'$ où $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \varepsilon_{i1}$, on peut exprimer la matrice de covariance de $\tilde{\varepsilon}_i$ « Σ_i » et son inverse « Σ_i^{-1} » en fonction de σ_ε^2 comme suit :

$$\begin{cases} \Sigma_i = \sigma_\varepsilon^2 (I_{m_i-1} + e_{m_i-1} e'_{m_i-1}) \\ \Sigma_i^{-1} = \sigma_\varepsilon^{-2} (I_{m_i-1} - e_{m_i-1} e'_{m_i-1} / m_i) \end{cases}$$

³⁹ Ces surfaces non-paramétriques sont issues de l'estimation de la fonction de lien lorsque le vecteur z est de dimension supérieure ou égale à 2. Dans la dernière section de ce chapitre nous décomposons l'inégalité éducative afin d'obtenir deux variables *within* et *between* qui sont toutes deux introduites dans le modèle où « $p=2$ ».

⁴⁰ Le modèle proposé par Henderson et al. (2008) présente le cas d'un panel cylindré alors que Zhou et Li (2011) ont généralisé la procédure itérative dans le cas de panel non cylindré.

Si on note $g_{it} = g(z_{it})$, le modèle (1) devient donc $y_{it} = g_{it} + u_i + \varepsilon_{it}$. Pour le cas où « $t=1$ » on a $y_{i1} = g_{i1} + u_i + \varepsilon_{i1}$.

Si on note de plus $\tilde{y}_i = (\tilde{y}_{i2}, \dots, \tilde{y}_{im_i})'$ et $g_i = (g_{i2}, \dots, g_{im_i})'$ où $\tilde{y}_{it} = y_{it} - y_{i1}$, on peut exprimer $\tilde{\varepsilon}_i$ en fonction de \tilde{y}_i , g_i et g_{i1} comme suit :

$$\begin{aligned} y_{it} &= g(z_{it}) + u_i + \varepsilon_{it} = g_{it} + u_i + \varepsilon_{it} \\ \Rightarrow \tilde{y}_{it} + y_{i1} &= g_{it} + u_i + \tilde{\varepsilon}_{it} + \varepsilon_{i1} \\ \Rightarrow \tilde{y}_{it} &= g_{it} - g_{i1} + \tilde{\varepsilon}_{it} \\ \Rightarrow \tilde{y}_i &= g_i - g_{i1} e_{m_i-1} + \tilde{\varepsilon}_i \\ \text{D'où } \tilde{\varepsilon}_i &= \tilde{y}_i - g_i + g_{i1} e_{m_i-1} \end{aligned}$$

A l'instar de Wang (2003), Lin et Carroll (2006) et Henderson et al. (2008), nous utilisons l'approche itérative basée sur la vraisemblance profilée appelée aussi « vraisemblance de profile ». Nous développons dans un premier temps, la dérivée première de la vraisemblance individuelle « $L_i(\cdot)$ » par rapport à g_{it} notée $L_{it}^g = \partial L_i(\cdot) / \partial g_{it}$. Ainsi on peut développer la vraisemblance individuelle à partir de la formulation de $\tilde{\varepsilon}_i$ comme suit :

$$\begin{aligned} L_i(\cdot) &= -\frac{1}{2} \tilde{\varepsilon}_i' \Sigma_i^{-1} \tilde{\varepsilon}_i, \quad i = 1, 2, \dots, n. \\ L_i(\cdot) &= -\frac{1}{2} (\tilde{y}_i - g_i + g_{i1} e_{m_i-1})' \Sigma_i^{-1} (\tilde{y}_i - g_i + g_{i1} e_{m_i-1}), \quad i = 1, 2, \dots, n. \\ L_{it}^g &= \frac{\partial L_i(\cdot)}{\partial g_{it}} = \begin{cases} -e_{m_i-1}' \Sigma_i^{-1} (\tilde{y}_i - g_i + g_{i1} e_{m_i-1}), & t = 1 \\ c_{i,t-1}' \Sigma_i^{-1} (\tilde{y}_i - g_i + g_{i1} e_{m_i-1}), & t \geq 2 \end{cases} \end{aligned}$$

où $c_{i,t-1}$ est un vecteur de dimension $(m_i - 1) \times 1$ dont tous les éléments sont nuls à l'exception du $(t-1)$ ème élément qui est égale à 1

$$\text{Si on définit } \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} g(z) \\ \partial g(z) / \partial z \end{pmatrix} = \begin{pmatrix} g(z) \\ g^{(1)}(z) \end{pmatrix} \text{ et } G_{it} = \begin{pmatrix} 1 \\ (z_{it} - z) / h \end{pmatrix},$$

l'estimation de $\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$ se fait en résolvant la condition du premier ordre de la vraisemblance profilée d'une manière itérative comme suit :

$$\sum_{i=1}^n \frac{1}{m_i} \sum_{t=1}^{m_i} K_h(z_{it} - z) G_{it} L_{it}^g \left(\hat{g}_{[l-1]}(z_{i1}), \dots, G_{it}(\alpha_0, \alpha_1)', \dots, \hat{g}_{[l-1]}(z_{im_i}) \right) = 0$$

où $\hat{g}_{[l-1]}(z_{is})$ est l'estimation de $g(z_{is})$ pour la $(l-1)$ ème itération et $k_h(v) = h^{-1} k(v/h)$ et $k(\cdot)$ est la fonction de kernel.

On peut par la suite définir l'estimation pour la l ème itération en fonction de la $(l-1)$ ème itération : $\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} \hat{g}_{[l]}(z) \\ \hat{g}_{[l]}^{(1)}(z) \end{pmatrix} = \frac{(A_1 + A_2)}{A_3}$ telle que :

$$\begin{cases} A_1 = \sum_{i=1}^n \frac{1}{m_i} \left(e'_{m_i-1} \Sigma_i^{-1} e_{m_i-1} K_h(z_{i1} - z) G_{i1} \hat{g}_{[l-1]}(z_{i1}) + \sum_{t=2}^{m_i} c'_{i,t-1} \Sigma_i^{-1} c_{i,t-1} K_h(z_{it} - z) G_{it} \hat{g}_{[l-1]}(z_{it}) \right) \\ A_2 = \sum_{i=1}^n \frac{1}{m_i} \left(-K_h(z_{i1} - z) G_{i1} e'_{m_i-1} \Sigma_i^{-1} H_{i,[l-1]} + \sum_{t=2}^{m_i} K_h(z_{it} - z) G_{it} c'_{i,t-1} \Sigma_i^{-1} H_{i,[l-1]} \right) \\ A_3 = \sum_{i=1}^n \frac{1}{m_i} \left(e'_{m_i-1} \Sigma_i^{-1} e_{m_i-1} K_h(z_{i1} - z) G_{i1} G'_{i1} + \sum_{t=2}^{m_i} c'_{i,t-1} \Sigma_i^{-1} c_{i,t-1} K_h(z_{it} - z) G_{it} G'_{it} \right) \end{cases}$$

où $H_{i,[l-1]}$ est un vecteur de dimension $(m_i - 1) \times 1$ dont les éléments notés « $h_{is,[l-1]}$ » sont tels que $h_{is,[l-1]} = (\tilde{y}_{it} - (\hat{g}_{[l-1]}(z_{it}) - \hat{g}_{[l-1]}(z_{i1})))$, $t = 1, 2, \dots, m_i$.

L'estimateur initial de $g(\cdot)$ est obtenu en se basant sur les séries temporelles tandis que la dernière itération est retenue dès lors que le critère de convergence est vérifié :

$$\frac{\sum_{i=1}^n \frac{1}{m_i} \sum_{t=2}^{m_i} (\hat{g}_{[l]}(z_{it}) - \hat{g}_{[l-1]}(z_{it}))^2}{\sum_{i=1}^n \frac{1}{m_i} \sum_{t=2}^{m_i} \hat{g}_{[l-1]}^2(z_{it})} \leq 0.01$$

Par ailleurs la variance σ_ε^2 est estimée par :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{2n} \sum_{i=1}^n \frac{1}{m_i - 1} \sum_{t=2}^{m_i} (y_{it} - y_{i1} - (\hat{g}_{[l]}(z_{it}) - \hat{g}_{[l-1]}(z_{it})))^2$$

La variance de l'estimateur $\hat{g}(z)$ est calculé par : $\kappa(nh\hat{\Omega}(z))^{-1}$

$$\text{où } \kappa = \int k^2(v) dv \text{ n et } \hat{\Omega}(z) = \sum_{i=1}^n \frac{m_i - 1}{m_i} \sum_{t=2}^{m_i} K_h(z_{i1} - z) / \hat{\sigma}_\varepsilon^2$$

Pour l'estimation du modèle semi-paramétrique, on définit l'estimateur non paramétrique des « q » variables de contrôle $\hat{g}_x(\cdot) = (\hat{g}_{x,1}(\cdot), \dots, \hat{g}_{x,q}(\cdot))'$ et de la variable dépendante $\hat{g}_y(\cdot)$ définit l'estimateur. Ainsi l'estimation de γ qui est de dimension $q \times 1$ est :

$$\hat{\gamma} = \left(\sum_{i=1}^n \frac{\tilde{x}'_i \Sigma_i^{-1} \tilde{x}_i}{m_i} \right)^{-1} \left(\sum_{i=1}^n \frac{\tilde{x}'_i \Sigma_i^{-1} \tilde{y}_i}{m_i} \right)$$

où \tilde{x}_i et \tilde{y}_i sont respectivement des matrices de dimension $(m_i - 1) \times q$ et $(m_i - 1) \times 1$ telles que la $s^{\text{ème}}$ ligne est définie par : $\tilde{x}_{is} = \tilde{x}_{is} - (\hat{g}_x(z_{is}) - \hat{g}_x(z_{i1}))$ et $\tilde{y}_{is} = \tilde{y}_{is} - (\hat{g}_y(z_{is}) - \hat{g}_y(z_{i1}))$. La composante non-paramétrique du modèle semi-paramétrique est déduite en remplaçant \tilde{y}_{it} par $\tilde{y}_{it} - \tilde{x}'_{it} \hat{\gamma}$.

2.4. L'impact des inégalités éducatives sur le développement économique : une vérification empirique de la non-linéarité

Comme nous l'avons mentionné dans la section 2.2, la majorité des travaux au niveau transnational⁴¹ n'ont pas pris en considération le caractère non linéaire de la relation. Dans le meilleur des cas, celle-ci a soit été évoquée et non vérifiée empiriquement, soit testée uniquement en introduisant une variable d'interaction (entre l'indice de Gini de l'éducation et la moyenne d'années de scolarisation ou entre l'indice de Gini de l'éducation et l'accroissement de la moyenne d'années de scolarisation).

Certains auteurs, comme Lopez et al. (1998), Park (2006) et Földvári et Van Leeuwen (2011), se sont basés sur le développement de Taylor d'ordre 2 de la fonction de production pour justifier d'un point de vue théorique la non linéarité de la relation. Ainsi, Lopez et al. (1998) montrent que le coefficient de variation de l'éducation affecte d'une façon non linéaire la fonction de production.

Berthélemy (2002) avance l'idée de la présence de clubs de convergence comme source de non linéarité dans laquelle l'échantillon retenu dans l'analyse comporte en même temps deux groupes de pays. Un premier groupe dont le niveau d'éducation est suffisamment élevé et un autre n'ayant pas encore atteint un seuil critique au-delà duquel peut se déclencher le processus de croissance. L'existence d'équilibres multiples justifie, selon l'auteur, l'existence d'une relation non linéaire.

Berthélemy et Arestoff (2003) évoquent aussi deux autres raisons qui alimentent cette source de non linéarité. La première fait référence au passage du niveau microéconomique au niveau macroéconomique tandis que la seconde concerne uniquement le niveau microéconomique. En effet, d'un côté, le passage du niveau individuel au niveau national renvoie au « problème d'agrégation » évoqué aussi par Thomas et al. (2001) et Földvári et Van Leeuwen (2011) comme une source importante de non linéarité de la relation⁴². De l'autre côté, même au niveau microéconomique, la relation est souvent considérée de nature non linéaire. Ainsi, l'introduction d'un terme quadratique du nombre d'années d'études et/ou d'expérience professionnelle est souvent utilisée en ce sens.

Sauer et Zagler (2014) ont introduit une variable multiplicative de la moyenne d'années de scolarisation et de l'indice de Gini de l'éducation et ont affirmé l'existence d'un lien non linéaire se manifestant à travers le niveau d'éducation. Pour les pays dont le niveau d'éducation est suffisamment faible, les auteurs trouvent un impact positif de l'inégalité d'éducation sur la croissance économique, tandis que l'impact devient négatif pour les pays ayant un niveau d'éducation moyen ou élevé.

D'autres auteurs, comme Bowman (2007), déjà évoqué à la section 2.2, et Castelló (2010b) qui n'ont pas explicitement évoqué la non linéarité de la relation,

⁴¹ Il faut souligner que contrairement aux travaux au niveau transnational, certains travaux au niveau infranational (Yu et al., 2015 ; Güngör, 2010) ont testé la non linéarité de la relation en introduisant des puissances d'ordre supérieur de l'indice de l'inégalité éducative.

⁴² Voir section 2.2 pour plus de détails sur ce problème.

ont cependant abouti à des conclusions qui corroborent celle-ci. Ainsi Bowman (2007) a introduit conjointement l'indice de Gini de l'éducation avec une variable d'interaction couplant l'accroissement du niveau de scolarisation à l'indice d'inégalité d'éducation. Les coefficients de ces deux variables sont statistiquement significatifs et de signes opposés. Celui de l'indice de Gini est négatif alors que celui de la variable d'interaction est positif.

D'autre part, Castelló (2010b) confirme un effet différencié selon le niveau de développement des pays en question. L'auteur témoigne d'un effet négatif pour les pays en développement et d'un effet différent selon les pays développés : l'impact est négatif pour les pays européens et paraît positif pour les pays anglo-saxons.

Toutefois, ces tentatives empiriques présentent des limites impactant significativement la robustesse des estimations. Bowman (2007) a uniquement utilisé 57 observations en coupe transversale. Sauer et Zagler (2014) et Castelló (2010b) ont utilisé une modélisation en GMM dynamique sur un large panel tout en aboutissant à des résultats différents. Si l'effet négatif est observé pour Sauer et Zagler (2014) pour les pays dont le niveau d'éducation est relativement élevé, il est plutôt décelé chez Castelló (2010b) pour les pays moins développés. Ceci peut être expliqué par le problème de corrélation entre l'indice de Gini de l'éducation et la moyenne d'année de scolarisation⁴³. L'introduction de ces deux variables dans un même modèle pose un réel problème de colinéarité. Celle-ci s'amplifie de manière substantielle si l'on introduit, dans le même modèle, un troisième terme d'interaction composé lui aussi de ces deux variables. De plus, l'utilisation d'une variable d'interaction impose une forme assez réductrice de la non linéarité de la relation et dont l'interprétation est sujette à ambiguïté.

Nous testons, dans ce qui suit, la non linéarité de la relation, en utilisant une modélisation non-paramétrique et semi-paramétrique inexplorée jusqu'à présent. Celle-ci bénéficie à la fois d'un lissage flexible des formes non-paramétriques et de la robustesse des estimations des données de panel. Dans une première étape, nous commençons par une modélisation paramétrique polynomiale qui nous aidera à identifier le comportement non linéaire global de la relation.

2.4.1. Modèle non linéaire polynomial

Afin de vérifier empiriquement l'existence d'une relation non linéaire entre inégalités éducatives et croissance économique, nous testons dans un premier temps un modèle paramétrique intégrant des puissances d'ordre supérieur au modèle de base {modèle (M1)}. Ceci revient à introduire une fonction polynomiale quadratique ou cubique de l'indice d'inégalité qui renseigne globalement sur le comportement de la relation selon différents niveaux d'inégalité.

Au lieu d'introduire une variable d'interaction, qui risque d'amplifier la colinéarité dans le modèle, nous proposons de tester un comportement non linéaire plus complexe qui se prête à une interprétation non ambiguë de la forme de la relation. La modélisation non linéaire polynomiale permet d'identifier d'éventuels

⁴³ Voir la section 1.4 sur chapitre I pour plus de détails sur la corrélation entre l'indice de Gini de l'éducation et la moyenne d'année de scolarisation.

changements de concavité en étudiant simplement les signes des termes linéaire, quadratique et cubique. Cette approche permet de saisir l'allure adéquate de la relation sur un large éventail de formes usuellement reconnues (courbe en U, en U inversé, en N, en N inversé...).

Soit le modèle polynomial de degré k suivant :

$$\left\{ \begin{array}{l} y_{it} = \sum_{k=0}^{K+1} \beta_k (IneqEdu_{it-5})^k + \gamma X_{it-5} + u_i + \varepsilon_{it} \\ \text{où } t = 1955, 1960, \dots, 2010; \quad i = 1, 2, \dots, 120 \end{array} \right. \quad (M1)$$

Nous considérons ci-dessus trois modèles paramétriques dérivés du modèle polynomial (M1) et que nous notons modèle (1), (2) et (3). Ces modèles concernent respectivement le modèle linéaire pris comme *benchmark* et les modèles polynomiaux quadratique et cubique.

$$y_{it} = \beta_0 + \beta_1 IneqEdu_{it-5} + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (1)$$

$$y_{it} = \beta_0 + \beta_1 IneqEdu_{it-5} + \beta_2 IneqEdu_{it-5}^2 + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (2)$$

$$y_{it} = \beta_0 + \beta_1 IneqEdu_{it-5} + \beta_2 IneqEdu_{it-5}^2 + \beta_3 IneqEdu_{it-5}^3 + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (3)$$

$$\text{où } X_{it-5} = (DepGov_{it-5}, Invest_{it-5}, Ouverture_{it-5}, PopGrowth_{it-5})'$$

$$t = 1955, 1960, \dots, 2010; \quad i = 1, 2, \dots, 120$$

Le tableau n° 11 présente les résultats de ces trois spécifications paramétriques appliquées à un panel mondial de 120 pays en développement entre 1955 et 2010. Ainsi, les colonnes (1,4,7 et 10) du tableau n° 11 montrent les résultats du modèle linéaire de référence. Le paramètre β_1 est négatif et significatif à 1% pour les quatre indices d'inégalité d'éducation. Le niveau agrégé d'éducation est supposé stimuler la croissance du PIB par habitant. Toutefois si l'éducation est mal répartie entre les individus, elle peut en constituer un frein. Ainsi, pour l'indice de Gini de l'éducation, une augmentation de 1% de cet indice, toutes choses étant égales par ailleurs, devrait diminuer de près de 0.19% le niveau du PIB par habitant. Les inégalités dans l'éducation compromettent les possibilités de s'instruire pour les populations défavorisées, limitant ainsi la mobilité sociale et le développement des compétences.

Les paramètres $\gamma_1, \gamma_2, \gamma_3$ et γ_4 correspondants respectivement aux quatre variables de contrôle (Dépenses_Gouv, Investissement, Ouverture et Pop_Growth) sont tous statistiquement significatifs. Ainsi, conformément aux enseignements théoriques, l'ouverture internationale a une incidence positive sur la croissance économique. L'élargissement du marché qu'elle permet favorise le transfert des technologies et la compétitivité de l'appareil productif. Les fortes croissances démographiques pèsent sur la croissance du PIB par habitant.

Le coefficient γ_2 , qui est significativement positif, montre que l'investissement stimule la croissance. De manière générale, l'investissement impacte à la fois l'offre et la demande de biens et services. D'une part il agit sur l'offre par l'accroissement du stock de capital fixe et permet ainsi de renforcer les

capacités productives et permettre souvent des gains de productivité. D'autre part, il agit sur la demande globale adressée aux entreprises.

Nos résultats laissent toutefois apparaître une relation négative entre les dépenses gouvernementales et le niveau du PIB par habitant. De manière générale, les dépenses publiques sont supposées être un facteur stimulant de la croissance, toutefois, les résultats empiriques divergent souvent (Devarajan, Swaroop et Zou, 1996; Easterly, Loayza et Montiel, 1997). L'impact dépendrait de la composition de ces dépenses mais aussi du degré d'efficacité de leur allocation.

Les colonnes (2, 5, 8 et 11) du tableau n° 11 présentent les résultats des modèles polynomiaux quadratiques des quatre indices d'inégalité {modèle (2)}. Les coefficients, tous significatifs, montrent l'existence d'une courbe en « U » puisque le terme quadratique est positif⁴⁴. Nous avons évalué le point de retournement. Il correspond à 0.82, 2.18, 2.19 et 0.37 respectivement pour l'indice de Gini de l'éducation, les indices d'entropie généralisée d'ordre 1 et 0.5 et l'indice d'Atkinson d'ordre 1. Etant donné que les points de retournement de ces quatre indices correspondent à des niveaux très inégalitaires de l'éducation, il apparaît que la courbe en « U » est plutôt tronquée à droite.

Ceci confirme une fois de plus que la tendance de la relation est largement en sens inverse et qu'elle peut révéler des coefficients significatifs pour les puissances supérieures à 2. Les colonnes (3, 6, 9 et 12) du tableau n° 11 présentent les résultats du modèle polynomial cubique {modèle (3)}. Les estimations des modèles ayant comme mesure de l'inégalité éducative, l'indice de Gini ou les indices d'entropie généralisée d'ordre 1 et 0.5, restent encore significatives pour la spécification cubique. Ce qui implique que, pour ces trois indices, la relation est susceptible de supporter un ajustement non linéaire encore plus poussé. Notons de plus que les termes linéaires et cubiques sont tous négatifs et significatifs confirmant une tendance non linéaire inverse entre inégalités éducatives et croissance, changeant de concavité selon le niveau d'inégalité.

L'estimation du modèle polynomial cubique de l'indice d'Atkinson (colonne 12 du tableau n° 11) donne un terme linéaire non significatif tandis que celle du modèle polynomial quadratique (colonne 11 du tableau n° 11) nous indique que les termes linéaire et quadratique sont tous les deux significatifs. Les résultats des estimations menées à travers l'indice d'Atkinson d'ordre 1 dévoilent l'existence d'une courbe en « U ». Ce qui signifie que, dans le cas de cet indice, lorsque le niveau d'inégalité est très élevé, une diminution de l'inégalité ne favorise pas particulièrement la croissance économique, du moins jusqu'à un certain seuil où l'effet s'inverse.

On constate de plus que pour les quatre indices d'inégalité, le pouvoir explicatif est plus important lorsque le modèle polynomial est de degré plus élevé. Ainsi le « R² Within » du modèle quadratique est supérieur à celui du modèle linéaire et le « R² Within » du modèle cubique est plus élevé de celui du modèle quadratique. Ce qui confirme une fois de plus que la courbe présente un comportement globalement non linéaire.

⁴⁴ Si le terme quadratique était négatif, la courbe de la relation aurait plutôt la forme d'une courbe en cloche.

Tableau 11: Estimation paramétrique à effets fixes du modèle linéaire, quadratique et cubique du panel de pays en développement

explicatives	Variable dépendante : Logarithme du PIB par habitant											
	Linéaire	Quadratique	Cubique	Linéaire	Quadratique	Cubique	Linéaire	Quadratique	Cubique	Linéaire	Quadratique	Cubique
	(1) ^a	(2) ^a	(3) ^a	(4) ^b	(5) ^b	(6) ^b	(7) ^c	(8) ^c	(9) ^c	(10) ^d	(11) ^d	(12) ^d
constante	8.468 (57.57)***	9.762 (57.04)***	10.176 (45.35)***	7.500 (52.55)***	7.949 (55.63)***	8.504 (60.47)***	7.748 (54.81)***	8.393 (59.87)***	8.978 (64.28)***	7.843 (50.18)***	8.990 (49.56)***	8.542 (41.11)***
Ineq_Edu	-1.886 (18.72)***	-6.367 (17.23)***	-9.110 (8.8)***	-0.374 (11.67)***	-1.086 (14.14)***	-2.498 (18.37)***	-0.414 (14.55)***	-1.314 (17.5)***	-2.858 (19.48)***	-2.309 (11.09)***	-10.303 (13.54)***	-2.975 (1.6)
Ineq_Edu ²	-	3.862 (12.53)***	9.276 (4.79)***	-	0.249 (10.1)***	1.342 (14.51)***	-	0.300 (12.79)***	1.524 (14.59)***	-	13.906 (10.87)***	-17.105 (2.34)**
Ineq_Edu ³	-	-	-3.233 (2.83)**	-	-	-0.221 (12.2)***	-	-	-0.265 (11.98)***	-	-	38.742 (4.3)***
DepGov	-0.009 (2.99)**	-0.007 (2.69)**	-0.007 (2.6)**	-0.008 (2.53)**	-0.009 (2.98)**	-0.010 (3.32)***	-0.009 (2.93)**	-0.010 (3.29)***	-0.008 (3.04)**	-0.003 (1.07)	-0.004 (1.18)	-0.004 (1.16)
Invest	0.007 (5.63)***	0.009 (7.61)***	0.009 (7.59)***	0.006 (4.04)***	0.007 (5.33)***	0.008 (6.21)***	0.006 (4.4)***	0.008 (6.24)***	0.008 (6.76)***	0.009 (5.89)***	0.009 (6.68)***	0.009 (6.62)***
Ouverture	0.003 (8.65)***	0.002 (5.83)***	0.002 (5.53)***	0.005 (11.04)***	0.004 (9.37)***	0.003 (7.57)***	0.004 (10.25)***	0.003 (7.78)***	0.002 (6.21)***	0.004 (9.89)***	0.003 (8.4)***	0.004 (8.53)***
PopGrowth	-0.087 (1.92)*	-0.022 (0.51)	-0.020 (0.47)	-0.164 (3.34)***	-0.113 (2.41)**	-0.061 (1.4)	-0.138 (2.9)**	-0.069 (1.56)	-0.025 (0.59)	-0.142 (2.86)**	-0.095 (2.02)**	-0.087 (1.87)*
Nb d'observations	1093	1093	1093	1093	1093	1093	1093	1093	1093	1093	1093	1093
Nb de pays	120	120	120	120	120	120	120	120	120	120	120	120
R ² Within	0.40	0.48	0.49	0.28	0.35	0.44	0.33	0.43	0.50	0.27	0.35	0.37

Source: Auteur.

Les chiffres entre parenthèses correspondent aux statistiques de Student ***, ** et * indiquent que les variables sont respectivement significatives à 1%, 5 % et 10%

^a Ineq_Edu correspond à l'indice de Gini de l'éducation

^b Ineq_Edu correspond à l'indice de Theil de l'éducation

^c Ineq_Edu correspond à l'entropie généralisée d'ordre 0.5 : GE(0.5)

^d Ineq_Edu correspond à l'indice Atkinson d'ordre 1 : A(1)

2.4.2. Modèles non-paramétrique et semi-paramétrique

Il est habituel de mener une modélisation non linéaire polynomiale avant d'entamer une modélisation non-paramétrique et semi-paramétrique. L'utilisation d'un modèle polynomial permet de saisir l'existence d'un comportement non linéaire de la relation, cependant certaines formes de non linéarité peuvent être ignorées en introduisant uniquement des termes quadratiques, cubiques ou d'ordre supérieur. En effet, l'utilisation d'une modélisation non-paramétrique ou semi-paramétrique est moins contraignante puisqu'elle n'exige pas au préalable une forme fonctionnelle donnée. Cette démarche permet d'ajuster au mieux le nuage de points de la relation tout en bénéficiant de fonctions non paramétriques lisses et souples au lieu de fonctions polynomiales prédéfinies à l'avance.

Nous proposons d'étudier la relation moyennant des modèles non-paramétriques et semi-paramétriques dans le cas des données de panel. Ainsi si on remplace la fonction polynomiale $\sum_{k=0}^{K+1} \beta_k (IneqEdu_{it-5})^k$ dans le modèle polynomial (M1) par la fonction $g(IneqEdu_{it-5})$ à estimer, le modèle devient un modèle semi-paramétrique. De plus, lorsqu'on élimine la partie paramétrique γX_{it-5} contenant les autres variables de contrôle, le modèle devient alors un modèle non-paramétrique. On note ainsi les modèles non-paramétrique (M2) et semi-paramétrique (M3) suivants :

$$\begin{cases} y_{it} = g(IneqEdu_{it-5}) + u_i + \varepsilon_{it} \\ \text{où } t = 1955, 1960, \dots, 2010; \quad i = 1, 2, \dots, 120 \end{cases} \quad (M2)$$

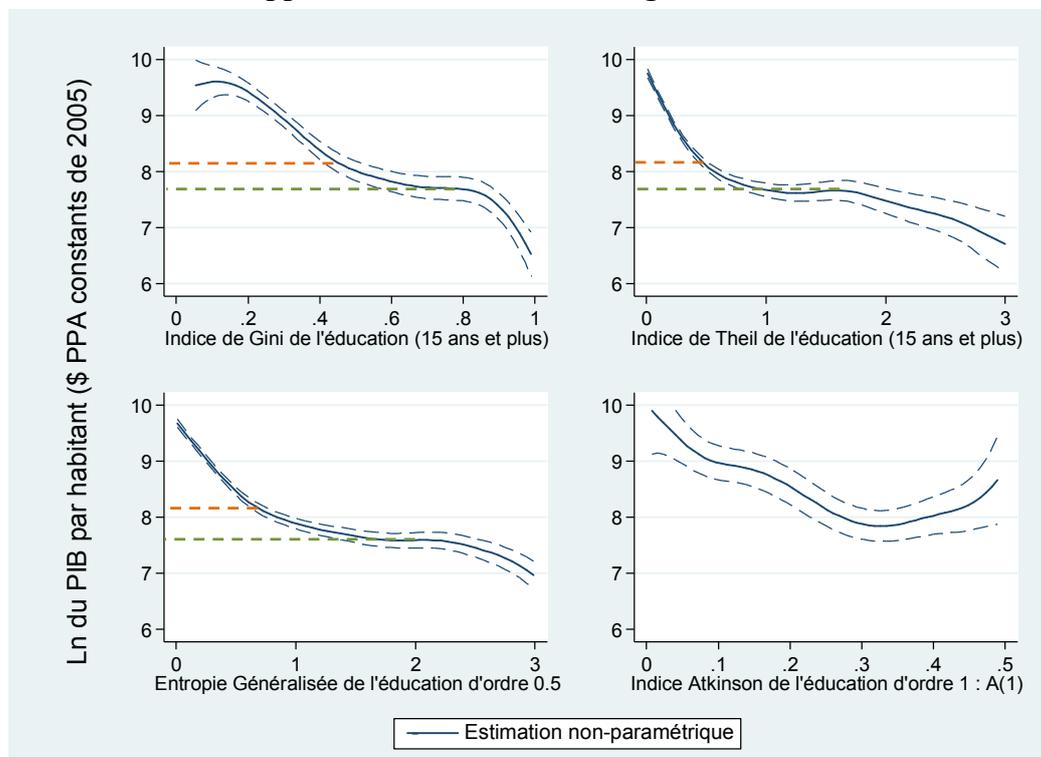
$$\begin{cases} y_{it} = g(IneqEdu_{it-5}) + \gamma X_{it-5} + u_i + \varepsilon_{it} \\ \text{où } t = 1955, 1960, \dots, 2010; \quad i = 1, 2, \dots, 120 \end{cases} \quad (M3)$$

Le graphique n°14 présente les résultats de l'estimation du modèle non-paramétrique (M2) pour les quatre indices d'inégalité retenus dans l'analyse. Ainsi, l'allure des courbes non-paramétriques obtenues s'avère être la même pour les indices de Gini et de l'entropie généralisée d'ordre 1 et 0.5 tandis qu'elle semble présenter une tendance différente pour l'indice d'Atkinson d'ordre 1.

En effet, concernant les trois premiers indices, on constate une relation globalement négative, celle-ci change de concavité tout en gardant cette même tendance à la baisse. Ainsi, lorsque le niveau d'inégalité éducative est suffisamment élevé {indice de Gini entre 0.80 et 1; indice de Theil entre 1.6 et 3; GE(0.5) entre 2 et 3}, la courbe non-paramétrique estimée présente une fonction décroissante concave. Celle-ci indique que l'effet de la réduction de l'inégalité éducative sur le développement économique est beaucoup plus important lorsque le niveau initial de l'inégalité d'éducation est plus élevé. Ceci confirme aussi que la pente de cette courbe est décroissante tout au long de la baisse de l'inégalité. Ainsi, la pente de la courbe diminue au fur et à mesure que l'indice de l'inégalité d'éducation baisse jusqu'à atteindre un premier seuil {0.80 pour l'indice de Gini, 1.6 pour l'indice de Theil et 2 pour GE(0.5)}. En deçà de ce niveau, la courbe non-paramétrique tend à suivre la même pente jusqu'à atteindre un second niveau d'inégalité {0.40 pour

l'indice de Gini, 0.80 pour l'indice de Theil et 1 pour GE(0.5)}, à partir de ce second seuil la fonction non-paramétrique commence à devenir convexe (la pente de la courbe devient croissante).

Graphique 14: Fonction non-paramétrique estimée pour le panel des pays en développement selon l'indice d'inégalité d'éducation



Source: Auteur.

L'analyse menée par rapport aux seuils d'inégalité d'éducation permet aussi de situer les niveaux de développement correspondants. Nous pouvons chercher à déceler ici les étapes que suit le développement au fur et à mesure de la diminution de l'inégalité éducative. Cette analyse ne concernera que les trois premiers indices étant donné que l'indice d'Atkinson ne suit pas la même tendance.

Le premier seuil du PIB par habitant, qui correspond à un indice de Gini de l'éducation de 0.8, est de 1883 \$ PPA constants de 2005. Pour les indices d'entropie généralisée d'ordre 1 et 0.5, ce premier seuil est évalué à 1850 et 1711 \$ PPA constants de 2005 respectivement. Nous avons donc presque le même seuil de PIB par habitant pour les trois indices d'inégalité éducative, soit un niveau de développement qui oscille aux alentours de 1800 \$ PPA constants de 2005. De même, le second seuil de développement se situe au voisinage de 3400 \$ PPA constants de 2005 pour les trois indices d'inégalité éducative.

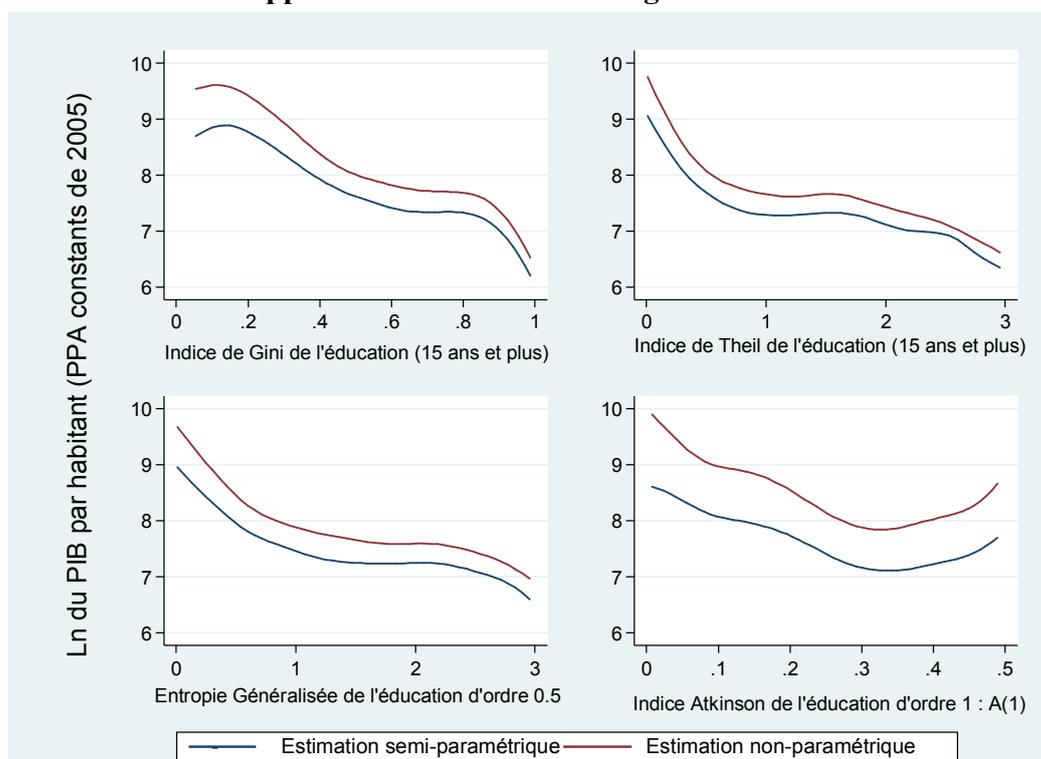
Ainsi, indépendamment de la mesure d'inégalité éducative utilisée, la réduction de l'inégalité d'éducation présente un impact plus prononcé sur le développement économique pour les pays à revenu faible (PIB par habitant inférieur à 1800 \$ PPA constants de 2005) ou les pays à revenu élevé (PIB par habitant supérieur à 3400 \$ PPA constants de 2005). A l'inverse, pour les pays à niveau de revenu intermédiaire (PIB par habitant qui se situe entre 1800 et 3400 \$

PPA constants de 2005), une réduction de l'inégalité s'avère présenter un impact moins important sur le développement économique.

Par ailleurs, la fonction non-paramétrique de l'indice d'Atkinson d'ordre 1, dévoile une tendance différente à celle mise en évidence pour les autres indices. Elle se présente sous la forme de « U » tronquée à droite. En effet, cette courbe montre une association négative pour les valeurs de l'indice d'Atkinson qui sont inférieures à 0.3 ($A_1 < 0.3$) et positive pour celles qui sont supérieures à 0.3 ($A_1 > 0.3$). Ainsi, pour cet indice, la réduction de l'inégalité éducative ne peut être bénéfique pour le développement économique que lorsque celle-ci est en deçà du seuil de 0.3 ($A_1 < 0.3$). Le niveau de développement correspond à ce seuil est de l'ordre de 2065 \$ PPA constants de 2005.

Afin de tester la robustesse de la fonction non-paramétrique estimée $g(\cdot)$, nous présentons une première comparaison entre les modèles semi-paramétriques ($M3$) et non-paramétriques ($M2$) pour vérifier si l'allure de la fonction estimée $g(\cdot)$ présente un comportement différent suite à l'introduction des variables de contrôle. Par la suite, nous testons aussi la robustesse des coefficients des variables de contrôle en comparant la partie paramétrique du modèle ($M3$) avec les résultats issus du modèle paramétrique polynomial ($M1$).

Graphique 15: Fonction semi-paramétrique pour le panel des pays en développement selon l'indice d'inégalité d'éducation



Source: Auteur.

Nous avons superposé les deux fonctions non-paramétrique {modèle ($M2$)} et semi-paramétrique {modèle ($M3$)} en un seul graphique (graphique n° 15) pour voir s'il existe, suite à l'introduction des variables de contrôle, un éventuel changement dans la tendance globale de la fonction estimée. Il apparaît que les

tendances des courbes semi-paramétriques sont identiques à une translation près à celles obtenues avec les modèles non-paramétriques. Ce qui indique que la relation demeure stable avec l'ajout de ces variables de contrôle.

Le tableau n° 12 présente les paramètres des variables de contrôle issus de l'estimation semi-paramétrique {modèle (M3)}. Les coefficients de ces variables présentent les mêmes signes qu'auparavant. Toutefois, le coefficient de l'accroissement de la population n'est plus significatif pour les trois premiers indices d'inégalité et celui de la part des dépenses gouvernementales ne l'est pas non plus pour l'indice d'Atkinson. Ces résultats rejoignent ceux trouvés dans la modélisation polynomiale. En effet, la croissance démographique est négative et significative pour le modèle linéaire tandis qu'elle devient non significative pour les modèles quadratique et cubique. Autrement-dit, la prise en compte de la non-linéarité de la variable d'inégalité d'éducation impacte la significativité de la variable de l'accroissement démographique. Ceci peut notamment être expliqué par un susceptible comportement non linéaire de cette variable.

Tableau 12: La partie paramétrique de la modélisation semi-paramétrique selon l'indice d'inégalité d'éducation utilisé

Variables explicatives	Variable dépendante : Logarithme du PIB par habitant			
	(1) ^a	(2) ^b	(3) ^c	(4) ^d
DepGov	-0.009 (2.70)**	-0.010 (3.19)***	-0.010 (3.13)**	-0.003 (0.80)
Invest	0.010 (6.40)***	0.009 (6.57)***	0.009 (6.36)***	0.009 (5.22)***
Ouverture	0.003 (5.00)***	0.003 (6.75)***	0.003 (6.75)***	0.004 (7.20)***
PopGrowth	-0.017 (0.34)	-0.024 (0.48)	-0.025 (0.49)	-0.083 (1.33)*
Nb d'observations	1093	1093	1093	1093
Nb de pays	120	24	24	24
R ² Within	0.73	0.74	0.74	0.72

Source: Auteur.

Les chiffres entre parenthèses correspondent aux statistiques de Student ***, ** et * indiquent que les variables sont respectivement significatives à 1%, 5 % et 10%

^a L'indice d'inégalité d'éducation correspond à l'indice de Gini de l'éducation

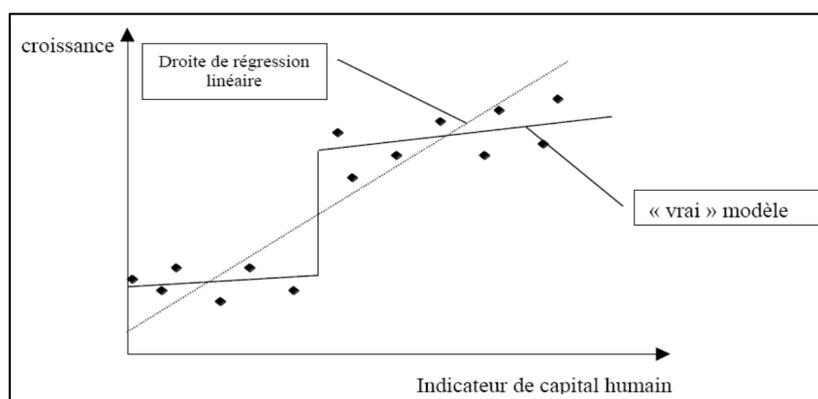
^b L'indice d'inégalité d'éducation correspond à l'indice de Theil de l'éducation

^c L'indice d'inégalité d'éducation correspond à l'entropie généralisée d'ordre 0.5 : GE(0.5)

^d L'indice d'inégalité d'éducation correspond à l'indice Atkinson d'ordre 1 : A(1)

Jusqu'à présent, toutes les estimations ont été menées sur un panel de 120 pays en développement. Vu l'hétérogénéité de ce panel, il serait intéressant de tester si la non linéarité persiste suite à une décomposition par zone géographique. Berthélemy et Arestoff (2003) justifient la non linéarité de la relation entre capital humain et croissance économique par la présence d'équilibres multiples (graphique n° 16) renvoyant à l'existence de clubs de convergence. L'hétérogénéité de l'échantillon des pays en développement peut être une cause de non linéarité.

Graphique 16: Illustration de la présence de clubs de convergence dans la relation éducation-croissance



Source : Berthélemy et Arestoff (2003)

Les pays en développement font l'objet, dans ce qui suit, d'une analyse par zone géographique⁴⁵. Les résultats concernant 24 pays développés sont aussi présentés en guise de comparaison. Ainsi nous donnons dans le graphique n° 17 neuf fonctions non-paramétriques établies en utilisant l'indice de Gini de l'éducation comme mesure de l'inégalité éducative. Les fonctions estimées correspondent respectivement aux pays du monde, pays développés, pays en développement, pays de l'Asie de l'Est et du Pacifique, pays de l'Europe et de l'Asie Centrale, pays de l'Amérique latine et des Caraïbes, pays de la zone MENA, pays de l'Asie du Sud et enfin les pays de l'Afrique subsaharienne.

La réduction de l'indice de Gini de l'éducation semble globalement contribuer à l'amélioration du niveau de développement économique dans l'ensemble des régions du monde. Cependant, l'ampleur de cet effet diffère d'une zone géographique à l'autre. A niveau d'inégalité éducative identique, une réduction de l'indice de Gini de l'éducation paraît être plus favorable pour les économies de l'Asie de l'Est et du Pacifique. A l'inverse, l'impact le moins important sur le développement économique s'observe pour les pays de l'Afrique subsaharienne. La fonction non-paramétrique de ces pays est presque horizontale indiquant que la croissance économique dans ces pays semble peu bénéficier de la diminution de l'inégalité éducative.

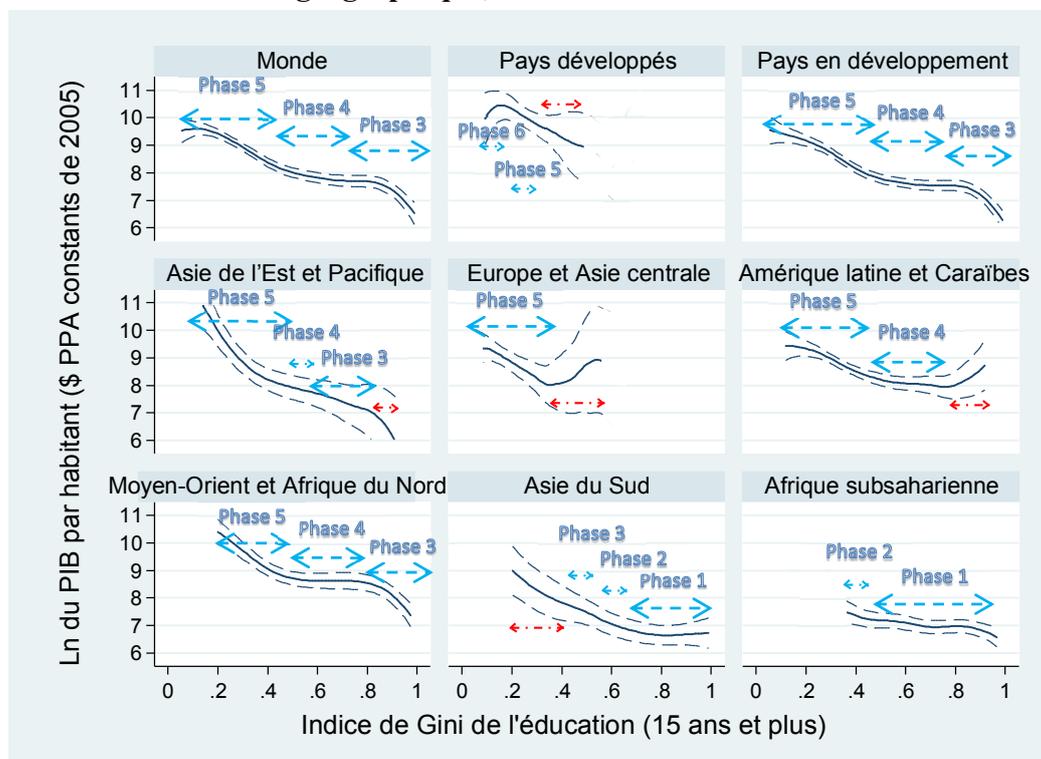
La fonction non-paramétrique de la région de l'Asie du Sud présente une pente nulle pour les valeurs du Gini d'éducation inférieures à 0.8. A partir de ce seuil, la pente est croissante au fur et à mesure que le niveau d'inégalité se réduit indiquant que pour les pays de cette région l'effet d'une réduction de l'inégalité éducative est plus bénéfique à la croissance économique lorsque son niveau initial est plus faible.

Les trajectoires des fonctions non-paramétriques des régions MENA et de l'Asie de l'Est ressemblent beaucoup à celle trouvée précédemment pour les pays en développement. Ainsi, l'effet de la réduction de l'inégalité éducative sur la croissance est plus marqué lorsque le niveau de l'inégalité d'éducation est soit assez élevé soit assez faible. Néanmoins, il faut souligner que, pour les valeurs de l'indice de Gini de l'éducation appartenant à l'intervalle [0.4 ; 0.8], la trajectoire des pays

⁴⁵ La décomposition géographique retenue est celle utilisée dans le chapitre I.

de l'Asie de l'Est fait exception et enregistre un cycle de croissance soutenue dans cet intervalle. A l'inverse, la fonction non-paramétrique de la région MENA est quasiment horizontale indiquant que la réduction de l'inégalité éducative dans l'intervalle [0.4 ; 0.8] est presque sans aucun effet sur le développement.

Graphique 17: Fonction non-paramétrique selon le niveau de développement et a zone géographique, Indice de Gini de l'éducation



Source: Auteur.

Ainsi, pour ces deux régions du monde, la diminution de l'inégalité d'éducation se montre plus bénéfique au développement économique lorsque le niveau initial de l'indice de Gini de l'éducation est soit très inégalitaire soit assez égalitaire. Pour des niveaux d'inégalité éducative intermédiaires, une réduction de celle-ci peut ne pas avoir d'impact, ou un impact relativement moins important dans le meilleur des cas.

Les régions de l'Amérique latine et Caraïbes, de l'Europe et de l'Asie centrale présentent des fonctions non-paramétriques différentes par rapport aux autres régions du monde. Celles-ci correspondent à une courbe en « U ». Néanmoins, la courbe de la région de l'Amérique latine est relativement plus aplatie. Soulignons que l'intervalle de confiance est particulièrement large à droite indiquant là que la relation est moins stable pour ces deux régions.

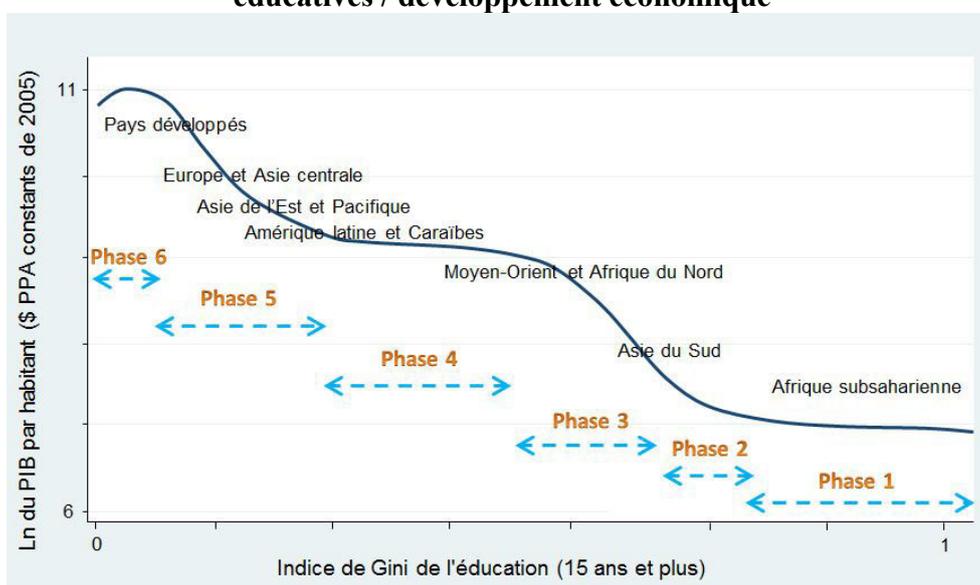
- **Les six phases de la relation**

Nous constatons que l'analyse menée précédemment par rapport aux seuils d'inégalité d'éducation ne permet pas, à elle seule, de mieux appréhender les phases successives que traverse la relation entre les inégalités d'éducation et le développement économique. C'est pour cette raison que nous essayons de proposer, dans ce qui suit, un schéma synthétique et diachronique qui résume les différentes

phases de cette relation. Nous nous appuyons sur les fonctions non-paramétriques des régions du monde mais aussi sur leurs intervalles de confiance. Ainsi, plus l'intervalle de confiance de la fonction est large sur une partie donnée de la trajectoire moins la relation est stable et par conséquent cette même partie ne sera pas prise en compte dans le schéma synthétique. Signalons que les zones des trajectoires relativement instables sont distinguées en rouge dans le graphique n° 17.

La question que nous allons donc essayer d'expliquer est pourquoi dans le même intervalle d'inégalité d'éducation, l'effet escompté d'une réduction de l'inégalité d'éducation se manifeste dans une région alors qu'il se fait attendre dans une autre ? Pour amorcer une réponse à cette question, nous allons conjuguer, à la fois, dans le schéma qui va suivre, les seuils d'inégalité d'éducation et ceux du développement (graphique n° 18).

Graphique 18: Schéma synthétique des phases de la relation inégalités éducatives / développement économique



Source: Auteur.

Remarquons que les trajectoires de l'Afrique subsaharienne sur l'intervalle $[0.4 ; 1]$ et de l'Asie du Sud sur l'intervalle $[0.8 ; 1]$, qui présentent un effet presque inexistant, se situent à un très faible niveau de PIB par habitant. En revanche, la trajectoire de la zone MENA sur l'intervalle $[0.8 ; 1]$, qui présente un impact très significatif, se positionne sur un niveau relativement moins faible de PIB par habitant. Ainsi on peut avancer que l'impact de la réduction de l'inégalité éducative sur le développement économique s'initie à partir d'un seuil critique de PIB par habitant. Si certains pays de l'Afrique subsaharienne restent en deçà de ce seuil critique de développement, les bénéfices économiques ne seront pas concrétisés même si l'indice de Gini de l'éducation passe de 1 à 0.4.

Ainsi la première phase de la relation est repérée principalement pour les pays de l'Afrique subsaharienne et quelques pays de l'Asie du Sud à certaines périodes. Cette phase est définie par rapport au niveau de développement et concerne les pays qui disposaient d'un niveau de PIB par habitant inférieur à 1100 \$ PPA constants de 2005.

La deuxième phase de la relation est caractérisée par un niveau de croissance relativement moyen et concerne essentiellement les pays de l'Asie du Sud et de l'Afrique subsaharienne. Elle peut être considérée comme transitoire à une phase de croissance soutenue. En revanche, la troisième phase, qui enregistre un niveau de croissance sensiblement meilleur que la deuxième, est plus observée pour les pays de la région MENA et certains pays de l'Asie de l'Est et du Pacifique. Ainsi la phase 2 comprend les pays dont le niveau de PIB par habitant se trouve compris entre 1100 et 1500 \$ PPA constants de 2005 tandis que les pays ayant franchi la phase 3 ont un niveau de PIB par habitant qui se situe entre 1500 et 3000 \$.

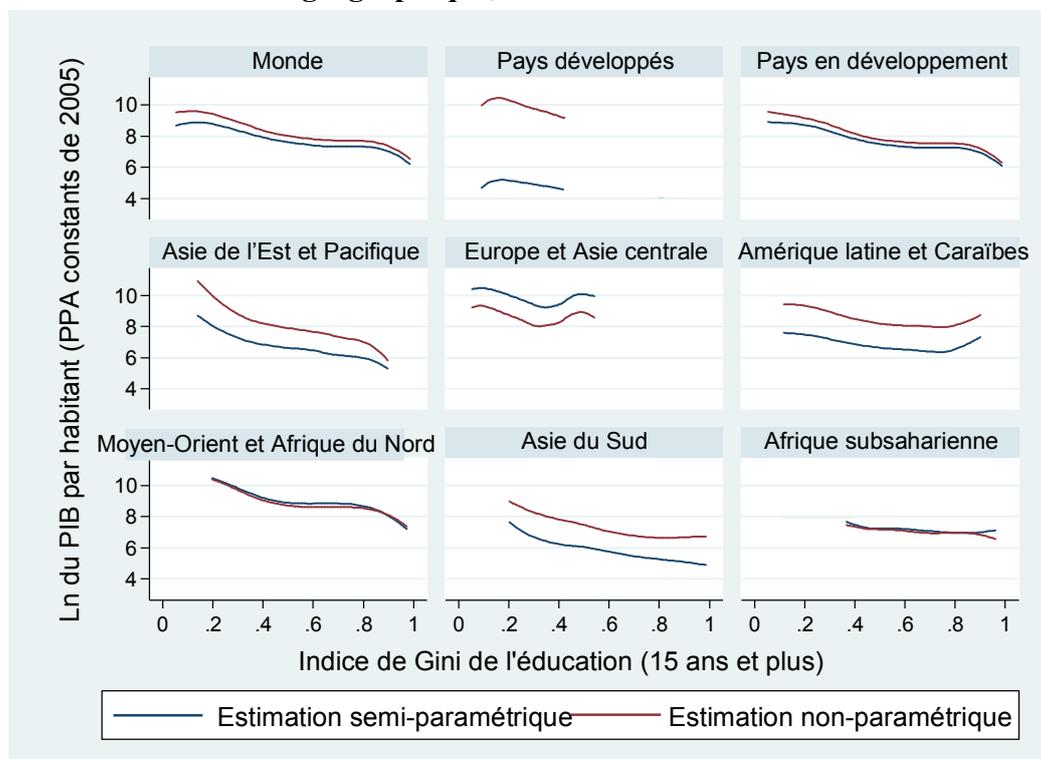
Quant aux phases 4 et 5 nous remarquons que chacune d'elles est définie par rapport à trois seuils. La quatrième phase est identifiée par deux seuils de niveau de développement et un autre relatif à l'inégalité éducative tandis que la cinquième est déterminée par deux seuils d'inégalité éducative et un autre par rapport au niveau de développement. Ainsi les pays qui ont un indice de Gini de l'éducation supérieur au seuil de 0.4 et dont le PIB par habitant est compris entre 3000 et 8000 \$ sont considérés comme étant situés à la phase 4. Les pays qui disposent d'un indice de Gini de l'éducation entre 0.2 et 0.4 et dont le PIB par habitant est supérieur à 8000 \$ sont considérés comme appartenant à la phase 5.

Contrairement aux cinq premières phases de la relation, la sixième phase n'apparaît plus prédominée par une tendance à la baisse de l'inégalité d'éducation. Elle est surtout marquée par des fluctuations de l'indice de Gini de l'éducation. Cette dernière phase, qui concerne principalement les pays développés, n'est caractérisée que par un indice de Gini de l'éducation inférieur au seuil de 0.2. La fonction non-paramétrique, dans cette phase, semble suivre une courbe en cloche tronquée à gauche indiquant là aussi que la relation n'est pas stable et qu'une augmentation de l'inégalité d'éducation pourrait aussi bien avoir un impact positif même sur la croissance économique. Ceci rejoint les résultats trouvés par Castelló (2010b) pour les pays développés et l'auteur soupçonne que le comportement de la relation peut même varier d'un pays développé à l'autre.

D'autres résultats sont aussi fournis dans les annexes 4.1, 4.2 et 4.3. Ils présentent les fonctions non-paramétriques selon le niveau de développement et la zone géographique pour les trois autres indices d'inégalité retenus dans l'analyse. Le schéma que nous avons proposé est vérifié si on utilise l'indice de Theil (Annexe 4.1) ou l'entropie généralisée d'ordre 0.5 (Annexe 4.2) et dans une moindre mesure pour l'indice Atkinson (Annexe 1.3). Ceci est notamment expliqué, comme on l'a vu précédemment, par le mode du calcul de cet indice qui diffère des autres.

Une comparaison transrégionale entre les modèles semi-paramétriques ($M3$) et non-paramétriques ($M2$) (graphique n° 19) permet d'examiner la stabilité des fonctions $g(\cdot)$ estimées suite à l'introduction des variables de contrôle. Les tendances des courbes semi-paramétriques sont une fois encore semblables à une translation près à celles issues à partir des modèles non-paramétriques. Ce qui montre que la relation reste inchangée après addition des variables de contrôle. De plus, l'effet global des variables de contrôle semble varier selon la région. En particulier, la contribution globale de ces variables illustrée par la différence entre les fonctions non-paramétriques et semi-paramétriques semble moins importante dans les pays du MENA et d'Afrique subsaharienne.

Graphique 19: Fonction semi-paramétrique selon le niveau de développement et la zone géographique, indice de Gini de l'éducation



Source: Auteur.

2.4.3. Décomposition inter et intra-groupes des inégalités et contribution au développement économique

Nous réalisons dans cette section deux types de décomposition⁴⁶ *within* et *between* permettant d'identifier le type d'inégalité éducative qui contribue le plus à la croissance du PIB par tête des pays en développement. Cette analyse consiste à intégrer dans la fonction non-paramétrique les deux composantes *within* et *between* de chacune des deux décompositions en tant que couple de variables.

En conséquence, la fonction non-paramétrique $g(\text{IneqEdu}_{it})$ de chacun des quatre indices d'inégalité devient donc une surface non-paramétrique notée $g^S(\text{IneqEdu}_{it-5}^{\text{Within}}, \text{IneqEdu}_{it-5}^{\text{Between}})$ où $\text{IneqEdu}_{it-5} = \text{IneqEdu}_{it-5}^{\text{Within}} + \text{IneqEdu}_{it-5}^{\text{Between}}$.

Cette surface non-paramétrique est estimée avec le même modèle que nous avons proposé précédemment. Dans ce cas, bivarié, z désigne un vecteur de dimension 2 tel que $z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ où $z_1 = \text{IneqEdu}_{it-5}^{\text{Within}}$ et $z_2 = \text{IneqEdu}_{it-5}^{\text{Between}}$. Nous présentons dans ce qui suit uniquement les estimations issues de la modélisation non-paramétrique puisqu'il s'est avéré que même en introduisant les variables de

⁴⁶ La méthodologie de la décomposition des indices d'inégalité éducative est décrite dans le chapitre I.

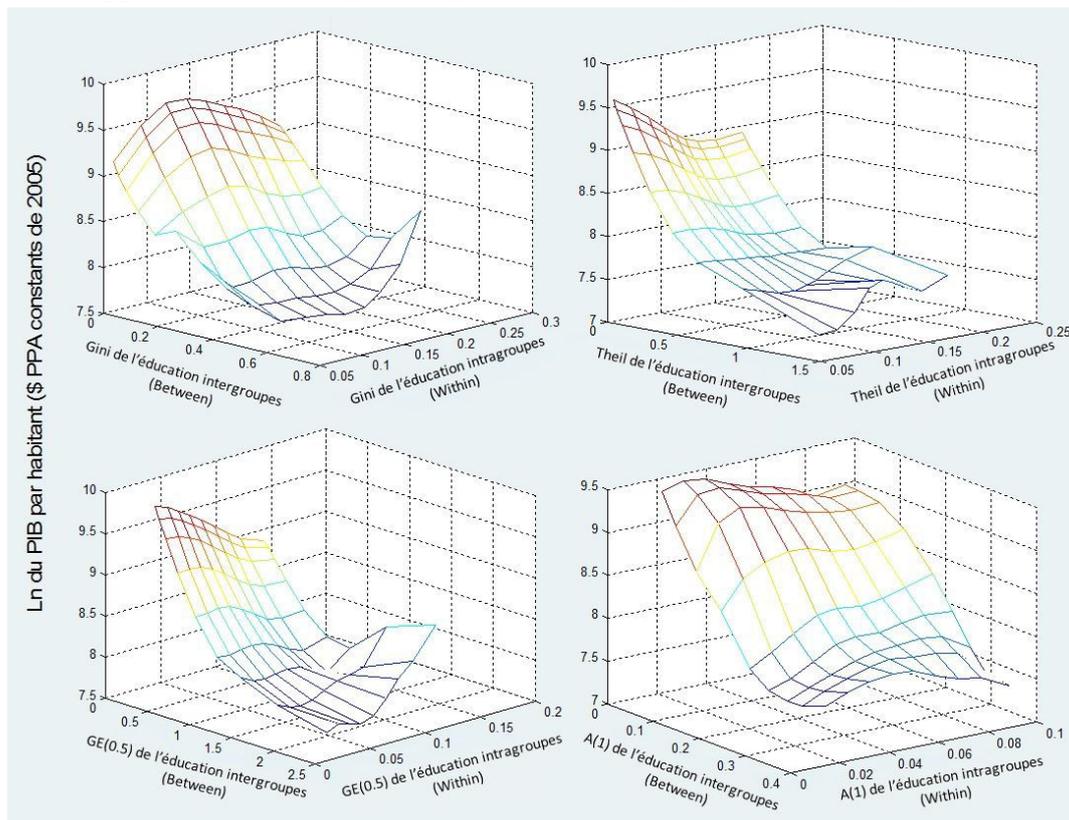
contrôle, l'allure de la surface semi-paramétrique reste très semblable à celle de la surface non-paramétrique.

La surface non-paramétrique est estimée à partir d'un ensemble de données qui fait correspondre à chaque couple d'inégalité d'éducation *within* et *between* ($IneqEdu_{it-5}^{Within}$, $IneqEdu_{it-5}^{Between}$) d'un pays "i" à l'année "t-5", son niveau de développement économique y_{it} . Ce modèle non-paramétrique est noté (M4) :

$$\begin{cases} y_{it} = g^S(IneqEdu_{it-5}^{Within}, IneqEdu_{it-5}^{Between}) + u_i + \varepsilon_{it} \\ \text{où } t = 1955, 1960, \dots, 2010; \quad i = 1, 2, \dots, 120 \end{cases} \quad (M4)$$

Nous retenons deux décompositions de l'inégalité d'éducation, la première concerne deux groupes (les illettrés et les non illettrés) et se décompose en une inégalité intra-éduqués et une inégalité inter-éduqués-illettrés et la seconde concerne 13 groupes (13 tranches d'âge allant de [15-19] à [75 et plus]) et se décompose en une inégalité intragénérationnelle et une inégalité intergénérationnelle.

Graphique 20. Décomposition de l'effet des inégalités éducatives sur le développement économique: inégalités intra-éduqués et inter éduqués-illettrés



Source: Auteur.

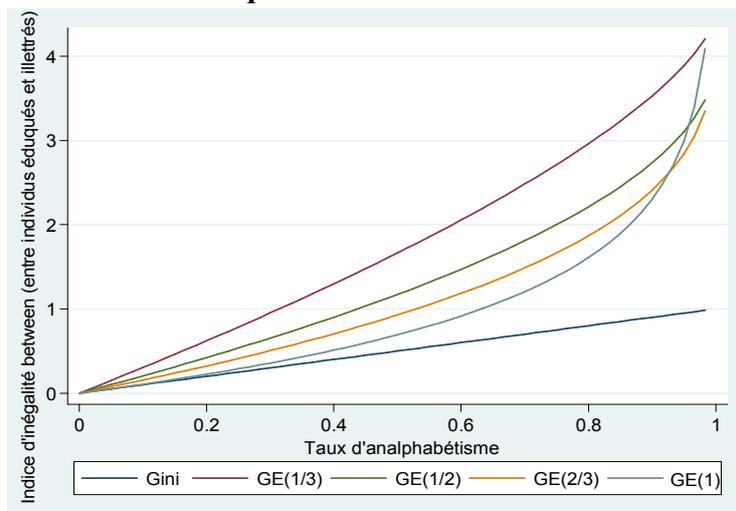
Notons "I" l'indice d'inégalité de l'éducation qui peut être un indice de Gini, Theil, d'entropie généralisée ou d'Atkinson. Pour tracer la surface 3D non-paramétrique de l'indice d'inégalité "I" de l'une des deux décompositions, nous nous sommes basés sur le programme d'estimation sous Gauss de la fonction $\hat{g}^S(I)$, pour programmer l'estimation de 81 couples correspondant aux différentes

combinaisons possibles des neuf déciles de la distribution de la composante *within* de l'indice I: $D_1^W(I)$ à $D_9^W(I)$ avec les neuf déciles de la distribution de la composante *between* de l'indice I: $D_1^B(I)$ à $D_9^B(I)$. Cette combinaison de couples nous permet de tracer en 3D le triplet (z_1, z_2, y) qui fait correspondre à chaque couple de déciles $(z_1, z_2) = (D_i^W(I), D_j^B(I))$: la fonction $\hat{g}^S(D_i^W(I), D_j^B(I))$ tel que $i, j = 1, 2, \dots, 9$.

Le graphique n° 20 représente les surfaces non-paramétriques permettant d'identifier l'effet des inégalités éducatives intra-éduqués et inter éduqués-illettrés sur le développement économique. Les surfaces des quatre indices d'inégalité dévoilent que l'augmentation du niveau du PIB par habitant des pays en développement est principalement due à la diminution des inégalités *between* (inter éduqués-illettrés).

Ce résultat peut être étendu aussi à la décomposition *les moins éduqués / les plus éduqués*. En effet, les surfaces non-paramétriques obtenues à partir de cette nouvelle décomposition fournissent les mêmes résultats trouvés avec l'autre décomposition *illettrés / éduqués*.

Graphique 21 Dépendance fonctionnelle liant le taux d'analphabétisme aux inégalités entre individus éduqués et illettrés selon différents indices d'inégalité

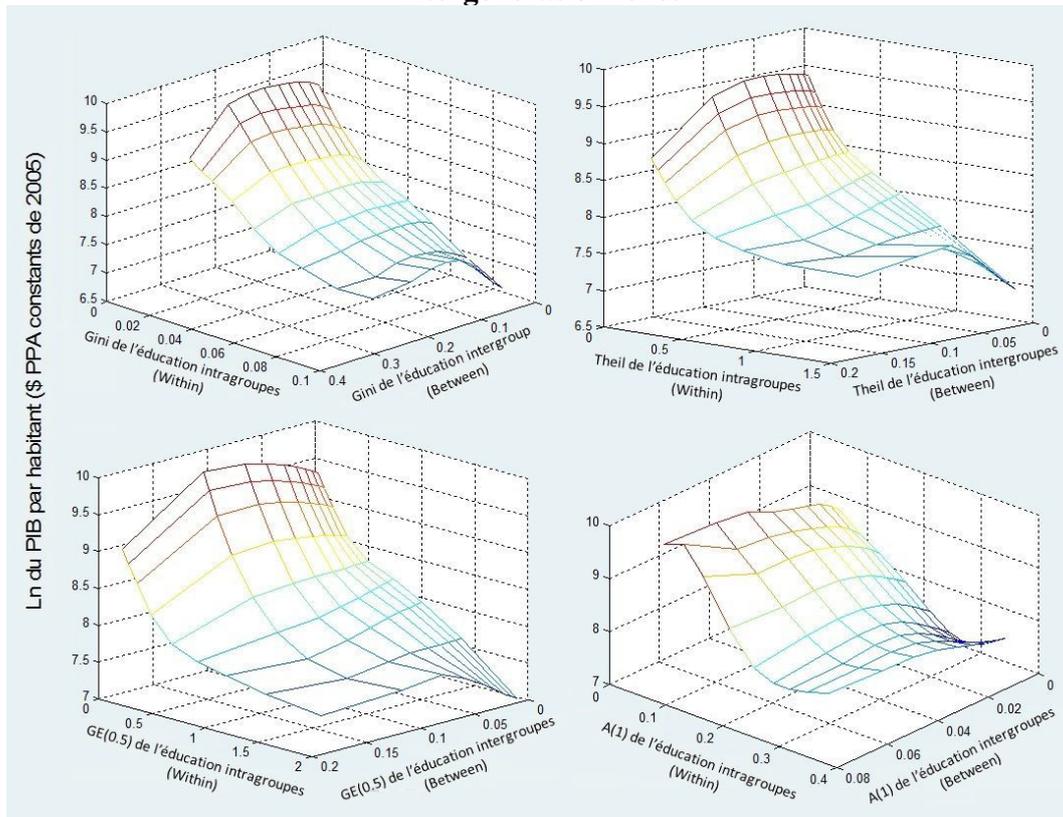


Source : Auteur.

Ainsi, la diminution des inégalités *between* (entre les plus et les moins éduqués) paraît stimuler la croissance économique plus que la réduction des inégalités au sein de la même catégorie (au sein des plus éduqués ou au sein des moins éduqués).

De plus, si on note « $\bar{\alpha}_{it}$ » le taux d'analphabétisme du pays « i » à l'année « t », nous avons montré, dans le chapitre I, que le taux d'analphabétisme représente l'inégalité entre les illettrés et les non illettrés (inégalité *between*) par rapport à l'indice de Gini de l'éducation $Gini_{it}^{Between} = \bar{\alpha}_{it}$.

Graphique 22 Décomposition de l'effet des inégalités éducatives sur le développement économique : inégalités intra-générationnelles et intergénérationnelles



Source: Auteur.

De même, par rapport à la famille des indices d'entropie généralisée, on peut aussi déduire que l'inégalité entre individus éduqués et illettrés dépend uniquement du taux d'analphabétisme⁴⁷. De plus, l'inégalité *between* est une fonction croissante du taux d'analphabétisme (voir graphique n° 21).

$$\begin{cases} Gini_{it}^{Between} = \bar{\alpha}_{it} \\ GE(r)_{it}^{Between} = \frac{[(1 - \bar{\alpha}_{it})^{1-r} - 1]}{r(r-1)} \quad \text{si } r \neq 1 \\ GE(r)_{it}^{Between} = -\ln(1 - \bar{\alpha}_{it}) \quad \text{si } r = 1 \end{cases}$$

Il apparaît par conséquent qu'il est plus bénéfique pour la croissance économique des pays en développement d'agir sur la moitié inférieure de la distribution de l'éducation, notamment en termes de réduction du taux d'analphabétisme et des sorties précoces du primaire. Ces dernières sont beaucoup plus liées aux inégalités de chance que des inégalités de compétence ou de mérite et ce sont celles qui favorisent le plus le développement.

Une deuxième décomposition est menée pour 13 générations (de 15-19 ans à 75 et plus). L'effet des inégalités éducatives intra-générationnelles et intergénérationnelles sur la croissance économique dans les pays en développement est représenté par les surfaces non-paramétriques du graphique n° 22. Contrairement à la première décomposition c'est plutôt la réduction des inégalités

⁴⁷ Voir la section 1.4.1

d'éducation à l'intérieur d'une même génération qui favorisent plus la croissance économique.

2.5. Conclusion

Ce chapitre a tenté de vérifier s'il existe une relation non linéaire entre le développement économique et les inégalités éducatives à partir d'un large panel de pays en développement entre 1955 et 2010. Pour cela, nous avons eu recours à des modèles non-paramétriques et semi-paramétriques en données de panel. Ces modèles très souples n'exigent pas au préalable une forme fonctionnelle donnée.

Nous avons montré d'un point de vue théorique l'importance d'introduire une variable qui puisse prendre en compte la distribution du capital humain dans les modèles de croissance. Cette variable est ignorée dans ces modèles car elle pose notamment le problème d'agrégation des capitaux humains individuels.

Nous avons dressé un bilan critique des travaux empiriques cherchant à saisir le rôle des inégalités d'éducation sur la croissance économique au niveau transnational. La discordance des conclusions peut être due aux choix des données mobilisées, au type de représentation des données (coupe transversale ou panel), aux mesures utilisées (absolues ou relatives), à l'échantillon retenu et à la non-linéarité de la relation.

Nous avons aussi expliqué le problème de colinéarité résultant de l'introduction de la moyenne d'années de scolarisation et de l'indice de Gini de l'éducation dans un même modèle. Ce problème est d'autant aggravé si l'on introduit, dans le même modèle, une variable d'interaction composée de ces deux variables. De là, nous avons choisi d'utiliser uniquement la variable de l'inégalité éducative.

Nous avons estimé dans un premier temps un modèle polynomial intégrant des puissances d'ordre supérieur à 1. Ce modèle a confirmé que la relation peut être non linéaire avec une tendance inverse entre les inégalités d'éducation et la croissance du PIB par tête, changeant de concavité selon le niveau d'inégalité.

Par la suite, l'estimation d'un modèle non-paramétrique pour les pays en développement nous a permis de déceler la nature de cette concavité. Ainsi lorsque la répartition de l'éducation est très inégalitaire la fonction est décroissante concave jusqu'à atteindre un premier seuil d'inégalité {0.80 pour l'indice de Gini, 1.6 pour l'indice de Theil et 2 pour GE(0.5)}. En deçà de ce niveau, la courbe non-paramétrique tend à suivre la même pente jusqu'à atteindre un second niveau d'inégalité {0.40 pour l'indice de Gini, 0.8 pour l'indice de Theil et 1 pour GE(0.5)}, à partir de ce second seuil la fonction non-paramétrique tend à devenir convexe (la pente de la courbe devient croissante).

Toutefois, cette relation semble présenter certaines spécificités. Une analyse par zone géographique dévoile des différences qui se manifestent par rapport à l'effet de la réduction de l'inégalité d'éducation. Ainsi, la courbe non-paramétrique des pays de l'Afrique subsaharienne est presque horizontale indiquant que la croissance économique dans ces pays ne semble pas atteindre le niveau de développement à partir duquel les effets bénéfiques de la diminution de l'inégalité éducative sur le processus de croissance soient perceptibles.

Nous avons proposé une analyse plus fine de la relation qui combine à la fois les seuils d'inégalité d'éducation et de développement pour étudier les trajectoires des régions du monde. Nous avons identifié six phases que traverse la relation où les trois premières sont définies seulement par rapport aux seuils de développement tandis que deux autres sont plutôt déterminées selon les niveaux d'inégalité d'éducation et de développement et une dernière phase se référant seulement à un seul seuil d'inégalité d'éducation.

Ainsi, la diminution de l'inégalité d'éducation présente le plus fort impact sur le développement économique lorsque le niveau initial de l'indice de Gini de l'éducation est soit très inégalitaire dépassant un premier seuil critique de développement (1500 \$ PPA constants de 2005), soit relativement égalitaire dépassant un second seuil critique de développement (8000 \$ PPA constants de 2005). Pour des niveaux d'inégalité éducative intermédiaires, une réduction de celle-ci peut ne pas avoir d'impact ou un impact relativement limité selon les cas. Les pays sont dans une situation où, peut-on dire, beaucoup d'autres facteurs entrent en ligne de compte. Néanmoins, d'une manière générale, la non linéarité de la relation est aussi bien confirmée pour les pays en développement que pour les grandes régions du monde retenues.

Nous nous sommes aussi intéressés, dans ce chapitre, à la question des arbitrages éducatifs liés à la distribution de l'éducation. Deux décompositions de l'inégalité ont été réalisées. La première concerne principalement les inégalités intra-éduqués et inter-éduqués-illettrés et la seconde concerne les inégalités intragénérationnelles et intergénérationnelles. Nos résultats montrent que pour stimuler la croissance dans les pays en développement, il est davantage bénéfique de réduire le taux d'analphabétisme que d'agir sur la moitié supérieure de la distribution de l'éducation. De même, la réduction des inégalités intragénérationnelles est plus importante pour les économies en développement que la réduction des inégalités intergénérationnelles.

La majorité des travaux, sur la question des arbitrages éducatifs, s'est exclusivement focalisée sur l'analyse des rendements privés de l'éducation selon le niveau d'enseignement au niveau microéconomique (Psacharopoulos et Patrinos, 2004 ; Colclough et al., 2010). Nous avons étudié deux types d'arbitrage, sur le plan macroéconomique, en poursuivant une démarche qui prend en considération le caractère non linéaire de la relation. Ainsi les pays en développement devraient favoriser le modèle de l'éducation de base plutôt que le modèle élitiste qui insiste sur l'extension de l'enseignement supérieur. De plus, ces pays devraient aussi se focaliser sur la diminution des inégalités éducatives au sein d'une même génération plutôt que de s'intéresser à réduire les inégalités éducatives entre les différentes générations.

Au-delà de la trajectoire globale de la non linéarité de la relation inégalités éducatives/croissance mise en évidence sur le plan transnational dans ce chapitre, un examen empirique sur des pays particuliers s'avère essentiel. Le chapitre III est dédié à l'étude du cas du Maroc.

Chapitre III

Les effets des inégalités éducatives sur le développement territorial : le cas du Maroc

- 3.1. Introduction**
- 3.2. Inégalités éducatives et développement humain dans les régions marocaines**
- 3.3. Inégalités éducatives et développement humain au niveau communal : une approche par l'économétrie spatiale**
- 3.4. Inégalités éducatives et développement économique des provinces marocaines**
- 3.5. Conclusion**

Chapitre III

Les effets des inégalités éducatives sur le développement territorial : le cas du Maroc

3.1. Introduction

Nous avons mis en évidence au chapitre II le caractère non linéaire de la relation entre inégalités éducatives et développement économique en analysant l'évolution des pays en développement. Il s'agit de vérifier dans ce chapitre si cette non-linéarité est aussi confirmée en analysant un pays particulier comme le Maroc.

Au-delà du schéma général mis en évidence sur le plan transnational dans le chapitre II, l'étude du Maroc, pour lequel nous disposons de données fines au plan régional, peut apporter un éclairage complémentaire sur la relation. En outre, à notre connaissance, aucune étude n'a traité les effets des inégalités éducatives sur le développement territorial au Maroc.

L'analyse de cette relation sur un ensemble de régions dans un pays réduit l'hétérogénéité rencontrée par rapport à un échantillon de pays et de ce point de vue on peut s'attendre à ce que cette non linéarité s'atténue ou soit peu prononcée.

A ce niveau géographique, il paraît intéressant d'explorer aussi de manière particulière les effets des inégalités éducatives sur le développement humain, en parallèle des effets sur le développement économique.

Nous retraçons dans une première section les principaux faits stylisés des inégalités éducatives et du développement humain au niveau des régions marocaines à travers l'estimation de l'indice de Gini de l'éducation et de l'indice de développement humain. Dans la section suivante, nous analysons les effets des inégalités éducatives sur le développement humain sur un ensemble de 1534 communes marocaines en 2004, à partir d'un modèle d'économétrie spatiale. Dans la dernière section, nous nous intéressons aux effets des inégalités d'éducation sur le développement économique en utilisant comme au chapitre II des modèles non-paramétriques et semi-paramétriques sur 61 provinces marocaines durant la période 1998-2013.

3.2. Inégalités éducatives et développement humain dans les régions marocaines

3.2.1. IDH et indice de Gini de l'éducation : revue de littérature et méthodologie

La désagrégation de l'IDH ou de l'indice de Gini de l'éducation permet de mettre en exergue les écarts en termes de développement humain et d'inégalités éducatives entre régions, provinces, communes, genre, ou milieux. Comme l'illustre le tableau n° 1, plusieurs tentatives d'estimation de ces deux indices ont été menées pour un certain nombre de pays à une échelle infra-nationale. Ainsi certains rapports nationaux du PNUD (RNDH) ont été consacrés à la désagrégation de l'IDH par province, régions ou districts, tout comme les rapports nationaux de

la Côte d'Ivoire (2013), de la république de Moldavie (2014), du Congo (2014), du Nepal (2014), de la Tanzanie (2014) ou de l'Ethiopie (2015) (cf. Tableau n° 13).

Tableau 13: Tableau synoptique des travaux concernant l'indice de Gini de l'éducation et de l'IDH au niveau infra-national

	Auteur	Cas étudié	Période	Echelle d'analyse	
Indice de Gini de l'éducation	Mesa (2007)	Philippines	1980-2000	78 provinces	
	Holsinger (2007)	Vietnam	1999	61 provinces	
	Loirel (2008)	Brésil	1950-2000	27 unités fédérales	
	Burt et Park (2009)	Corée du Sud	1970-2000	16 provinces & villes métropol.	
	Tomul (2009)	Turquie	1975, 2000	26 régions	
	Morales et Teran (2010)	Argentine	2002, 2007	32 régions	
	Fidalgo et al. (2010)	Portugal	1986-2005	20 régions	
	Kumba (2010)	Indonésie	1999-2005	23 provinces	
	Gustave et Joubert (2012)	Cameroun	2001, 2007	10 régions	
	Senadza (2012)	Ghana	2005	10 régions	
	Trabelsi (2013)	Tunisie	1975- 2004	24 gouvernorats	
	Yang et al. (2014)	Chine	1996- 2008	31 provinces	
	Agrawal (2014)	Inde	1993- 2009	18 grands Etats	
Indice de développement humain	Articles	Lipták (2010)	Roumanie et Hongrie	2004, 2007	8 régions
		Silva et Lopes (2014)	Portugal	2004, 2008	30 régions
		Schrott et al. (2015)	Autriche	2008	121 districts
		Hazell et al. (2012)	Canada	2000-2011	13 provinces
		Sabermahani et al. (2013)	Iran	2001, 2009	28 provinces
		Felice et Vasta (2015)	Italie	1871-2007	20 régions
	RNDH	Côte d'Ivoire, 2013	Côte d'Ivoire	2002-2011	11 régions
		Moldova, 2014	Moldavie	2013	4 régions
		Congo, 2014	Congo	2012	11 provinces
		Nepal, 2014	Nepal	2011	75 districts
Tanzanie, 2014	Tanzanie	2012	21 régions parmi les 26		
Ethiopie, 2015	Ethiopie	2005, 2013	11 régions & villes-régions		

Source : Auteur.

En recourant à la nouvelle méthodologie du PNUD certains auteurs ont calculé l'IDH au niveau infra-national comme c'est le cas de Hazell et al. (2012), pour 13 provinces du Canada durant la période 2000-2011; de Schrott et al. (2015), pour 121 districts de l'Autriche et de Felice et Vasta (2015), pour 20 régions italiennes pour la période 1871-2007, tandis que d'autres ont eu recours à l'ancienne méthodologie comme c'est le cas de Sabermahani et al. (2013), pour 28 provinces de l'Iran en 2001 et 2009 ; de Lipták (2010), pour 8 régions de la Roumanie et de la Hongrie en 2004 et 2007 (cf. Tableau n° 1).

D'autres auteurs ont aussi tenté d'appréhender les inégalités d'éducation au niveau infra-national à travers l'estimation de l'indice de Gini de l'éducation (Mesa, 2007; Holsinger, 2007; Loirel, 2008; Burt et Park, 2009; Tomul, 2009; Morales et Teran, 2010; Kumba, 2010; Gustave. et Joubert, 2012; Senadza, 2012; Trabelsi, 2013; Yang et al., 2014 et Agrawal, 2014). Ces auteurs ont analysé respectivement le cas des Philippines, Vietnam, Brésil, Corée du Sud, Turquie, Argentine, Indonésie, Cameroun, Ghana, Tunisie, Chine et l'Inde (cf. Tableau n° 1).

Toutefois, comme on l'a souligné dans le chapitre I, les résultats que l'on pourrait obtenir dépendent des choix méthodologiques retenus. En effet, certains travaux qui concerne l'IDH ou l'indice de Gini de l'éducation se heurtent à des difficultés méthodologiques liées principalement à la non disponibilité de l'information au niveau infra-national et au problème de mesure de certains indicateurs. Un de ces principaux problèmes relevé à la fois pour ces deux indicateurs est le recours à un nombre assez réduit de niveaux d'éducation pour

évaluer l'indice de Gini de l'éducation ou la composante "éducation" de l'IDH, ce qui implique une sous-estimation de l'indice d'inégalité et une surestimation de la composante éducation de l'IDH.

Morales et Teran (2010) confirment ce constat pour les travaux analysant l'indice de Gini de l'éducation et rapportent que "Some precision is lost in working with aggregate data because the reduced number of categories of level of education variable tends to underestimate the results".

L'idéal serait de disposer des données désagrégées par niveau d'éducation au sein de chaque cycle d'enseignement. Encore faut-il avoir cette répartition par niveau d'enseignement pour la même population et ce pour l'ensemble des pays, régions, provinces ou communes de l'échantillon. Potancokova et al. (2014) mentionnent la difficulté de disposer de telles données pour un panel de pays particulièrement pour les pays arabes et rapportent "Finding a sufficient number of countries with detailed data on education by both the level and grade completed was challenging for some regions because data are mostly collected for the highest completed level and not for information on completed grades. While the coverage was rather good for Latin America, Asia and Sub-Saharan Africa, finding data for Arab countries was much more complicated."

Toutefois, s'il apparaît difficile d'en disposer au niveau national, l'exercice est encore plus périlleux au niveau infra-national. En effet, la quasi-totalité des travaux traitant le niveau infra-national utilisent des données agrégées par cycle d'enseignement pour évaluer l'inégalité puisqu'une ventilation détaillée n'est souvent pas disponible. Dans ce cas, l'inégalité estimée concerne uniquement l'inégalité inter-cycles alors que l'inégalité intra-cycles est de ce fait ignorée.

De même, l'estimation de la moyenne des années de scolarisation dans la majorité des travaux sur l'IDH désagrégé au niveau infra-national est biaisée pour la même raison. On relève aussi dans ces travaux une confusion entre les personnes qui n'ont pas achevé un cycle d'enseignement et celles qui l'ont accompli (Hazell et al., 2012, utilisent par exemple uniquement 3 niveaux d'éducation). Ces travaux attribuent donc de facto la durée totale de scolarisation dans un cycle donné à des individus qui ne l'ont pas forcément terminé, ce qui est de nature à surestimer la moyenne des années de scolarisation (Tableau n° 2).

Pour remédier à cette insuffisance, certaines études proposent d'attribuer la moitié de la durée totale du cycle d'enseignement à tous ceux qui n'ont pas encore achevé ce dernier. Certes, cette approche permet de limiter en quelque sorte le biais d'estimation par rapport à celle qui affecte la durée totale à l'ensemble des individus. Toutefois, cette dernière attribue d'une manière arbitraire la moitié de la durée totale du cycle d'enseignement à tous les individus qui n'ont pas achevé ce cycle sans aucune distinction, ce qui est de nature à biaiser l'estimation. Jorda (2015) rapporte que "The main shortcoming of these studies is the assumption that all individuals who have not completed a particular educational level have the same arbitrary number of years of schooling. This choice introduces some measurement error, but the main shortcoming is that it is not possible to determine the direction of the bias. If our choice lie above (below) the actual mean of the incomplete level, the mean years of schooling of the educational cycle would be overestimated (underestimated)" et explique que le choix de cette démarche est encore plus

problématique puisqu'on ne peut affirmer si ce biais va dans le sens de la surestimation ou la sous-estimation de la moyenne d'années de scolarisation.

Tableau 14: Nombre de niveaux d'éducation et population d'intérêt retenus dans quelques travaux concernant l'IDH et le Gini d'éducation au niveau infra-national

	Auteur	Cas étudié	Nombre de niveaux d'éducation retenu	Population d'intérêt
Indice de Gini de l'éducation	Mesa (2007)	Philippines	7	15 ans et plus
	Holsinger (2007)	Vietnam	5	15 ans et plus
	Lorel (2008)	Brésil	6	5 ans et plus
	Burt et Park (2009)	Corée du Sud	9	6 ans et plus
	Tomul (2009)	Turquie	6	25 ans et plus
	Kumba (2010)	Indonésie	6	15 ans et plus
	Gustave et Joubert (2012)	Cameroun	7	15 ans et plus
	Senadza (2012)	Ghana	9	15 ans et plus
	Trabelsi (2013)	Tunisie	5	10 ans et plus
	Yang et al. (2014)	Chine	5	6 ans et plus
	Agrawal (2014)	Inde	7	15 ans et plus
IDH	Hazell et al. (2012)	Canada	3	25 ans et plus
	Schrott et al. (2015)	Autriche	5	25-64 ans
	Felice et Vasta (2015)	Italie	5	6 ans et plus
	RNDH Ethiopie (2015)	Ethiopie	6	25 ans et plus

Source : Auteur.

Un autre problème méthodologique dans cette littérature réside dans le choix de la population sur laquelle aussi bien la moyenne d'années de scolarisation que l'indice d'inégalité éducative sont évalués. Il est retenu le plus souvent la population âgée de 15 ans et plus pour les travaux analysant l'indice de Gini de l'éducation et la population âgée de 25 ans et plus pour les travaux traitant l'IDH. Toutefois, comme le montre le tableau n° 14, l'absence de données pour ces tranches d'âge de population communément utilisées contraint certains travaux à mener des analyses en se référant à des populations peu représentatives. A titre d'exemple, des travaux analysant l'indice de Gini de l'éducation comme Lorel (2008), Burt et Park (2009), Trabelsi (2013) et Yang et al. (2014) utilisent, selon les populations âgées de 5 ans et plus, 6 ans et plus, 10 ans et plus. Le constat est le même pour les travaux qui traitent l'IDH. Par exemple, Felice et Vasta (2015) utilisent la population âgée de 6 ans et plus.

Le choix de la population sur laquelle on souhaite évaluer l'indice de Gini de l'éducation ou la composante "éducation" de l'IDH impacte sensiblement les résultats obtenus. Ce biais est principalement dû à l'intégration dans le calcul des indices⁴⁸ de la population âgée de moins de 15 ans majoritairement en cours de scolarisation comme étant une population qui a achevé sa scolarisation.

C'est pour cette raison que la mesure de la dimension "éducation" de l'IDH s'est limitée dans certain cas à l'utilisation de la moyenne des années de scolarisation exclusivement et ignore par conséquent la durée de scolarisation espérée des jeunes entre 5 et 24 ans poursuivant leur scolarité (Schrott et al., 2015 ; Silva et Lopes, 2014). Même lorsqu'elle est réalisée, l'estimation de l'espérance de vie scolaire est moins précise puisqu'elle ne s'appuie dans le meilleur des cas que sur quatre taux spécifiques de scolarisation pour des tranches d'âge agrégées

⁴⁸ La moyenne d'années de scolarisation ou l'indice de Gini de l'éducation.

correspondant aux quatre cycles d'enseignement (voir par exemple Felice et Vasta, 2015). En effet, les taux de scolarisation dans ces travaux sont alors considérés uniformes pour les différents âges.

Ainsi pour remédier à ces insuffisances méthodologiques et contrairement aux anciens travaux, nos estimations de l'espérance de vie scolaire⁴⁹ sont basées sur un ensemble de 20 taux spécifiques de scolarisation par âge simple⁵⁰ calculés à partir d'une compilation de différentes sources statistiques (recensements scolaires dans les deux secteurs public et privé, données universitaires, ENDPR). De même, nos estimations de la moyenne d'années de scolarisation⁵¹ et de l'indice de Gini de l'éducation sont aussi basées sur un ensemble de 22 niveaux d'enseignement à partir de données désagrégées issues des recensements de la population marocaine (1982, 1994 et 2004) et de l'ENDPR 2009-2010. Ces derniers donnent, de manière plus exacte, le niveau maximum de scolarisation atteint par les marocains âgés de 15 ans et plus/ 25 ans et plus dans tous les niveaux d'enseignement (voir tableau n°3).

Le tableau n° 15 présente le nombre d'années de scolarisation correspondant à chacun des niveaux d'éducation désagrégés retenus dans nos estimations selon l'ancienne et la nouvelle structure du système éducatif marocain⁵². Au total 22 niveaux d'enseignement sont considérés (tableau n° 3) allant de la première année du cycle primaire jusqu'à la troisième année de doctorat (y compris deux autres niveaux, noté préscolaire et illettré). Ces données sont réparties selon le découpage administratif du territoire marocain (région, province ou commune) et ventilés par genre et par milieu (rural ou urbain).

L'espérance de vie scolaire (EVS) est définie par l'UNESCO comme étant le nombre total d'années de scolarité qu'un enfant d'un certain âge peut s'attendre à recevoir dans l'avenir, en supposant que la probabilité de son admission à l'école à un âge donné est égale au taux de scolarisation actuel de cet âge. A des fins de comparaison internationale, nous avons utilisé la méthode de l'UNESCO qui consiste à sommer l'ensemble des taux spécifiques par âge simple. Le taux spécifique à un certain âge "a" est calculé en rapportant la somme des effectifs scolarisés dans les différents cycles d'enseignement ayant l'âge "a" sur la totalité de la population ayant l'âge "a".

$$TS_a = \left(\frac{E_{\text{prim},a} + E_{\text{col},a} + E_{\text{sec},a} + E_{\text{sup},a}}{POP_a} \right)$$

avec TS_a : Le taux spécifique à l'âge « a »

$E_{\text{prim},a}$: Effectif des scolarisés dans le primaire qui ont l'âge « a »

$E_{\text{COL},a}$: Effectif des scolarisés dans le secondaire collégial qui ont l'âge « a »

$E_{\text{LYC},a}$: Effectif des scolarisés dans le secondaire qualifiant qui ont l'âge « a »

⁴⁹ Nous avons utilisé la nouvelle méthodologie du PNUD où l'espérance de vie scolaire et la moyenne d'années de scolarisation de la population âgée de 25 ans représentent les deux sous-dimensions de la composante "éducation" de l'IDH.

⁵⁰ Au total 20 taux spécifiques sont considérés allant du taux spécifique à l'âge de 5 ans à celui de 24 ans.

⁵¹ La moyenne d'années de scolarisation (MAS) est retrouvée dans les travaux concernant aussi bien le Gini d'éducation que l'IDH. Dans le premier type de travaux, c'est la MAS de la population âgée de 15 ans et plus qui est considérée tandis que dans le deuxième type de travaux c'est plutôt la MAS de la population âgée de 25 ans et plus qui est utilisée comme seconde sous-dimension de la composante "éducation" de l'IDH dans la nouvelle méthodologie du PNUD (à partir de l'année 2010).

⁵² L'ancienne structure du système éducatif est retrouvée dans les données de l'année 1982 où le cycle primaire et collégial comportait 5 et 4 années d'enseignement respectivement.

$E_{sup, a}$: Effectif des scolarisés dans le supérieur qui ont l'âge « a »
 POP_a : Effectif de la population qui a l'âge « a ».

Ainsi l'EVS est la somme des 20 taux spécifiques par âge simple (de l'âge de 5 ans à l'âge de 26 ans)

$$EVS = \sum_{a=5}^{26} TS_a$$

Tableau 15: Nombre d'années de scolarisation correspondant à chaque niveau d'éducation selon l'ancienne et la nouvelle structure du système éducatif marocain

Nouvelle structure du système éducatif marocain		Ancienne structure du système éducatif marocain		"y _i " : Nombre d'années de scolarisation
Cycle	Niveau d'éducation	Cycle	Niveau d'éducation	
1	Néant	Illettré	Illettré	0
2	Préscolaire	Préscolaire	Préscolaire	1
3	Primaire	1 ^{ère} année primaire	Primaire	1 ^{ère} année primaire
4		2 ^{ème} année primaire		2 ^{ème} année primaire
5		3 ^{ème} année primaire		3 ^{ème} année primaire
6		4 ^{ème} année primaire		4 ^{ème} année primaire
7		5 ^{ème} année primaire		5 ^{ème} année primaire
8	6 ^{ème} année primaire	Collégial	1 ^{ère} année collégiale	6
9	Secondaire collégial		2 ^{ème} année collégiale	7
10			3 ^{ème} année collégiale	8
11			4 ^{ème} année collégiale	9
12	Secondaire qualifiant	Tronc commun	Lycée	1 ^{ère} année du lycée
13		1 ^{ère} année du baccalauréat		2 ^{ème} année du lycée
14		2 ^{ème} année du baccalauréat		3 ^{ème} année du lycée
15	Supérieur	1 ^{ère} année supérieur	Supérieur	1 ^{ère} année supérieur
16		2 ^{ème} année supérieur		2 ^{ème} année supérieur
17		3 ^{ème} année supérieur		3 ^{ème} année supérieur
18		4 ^{ème} année supérieur		4 ^{ème} année supérieur
19		5 ^{ème} année supérieur		5 ^{ème} année supérieur
20		6 ^{ème} année supérieur		6 ^{ème} année supérieur
21		7 ^{ème} année supérieur		7 ^{ème} année supérieur
22		8 ^{ème} année supérieur		8 ^{ème} année supérieur

Source : Auteur.

La moyenne d'années de scolarisation (MAS) est calculée en utilisant 22 niveaux à partir d'une moyenne pondérée entre les proportions de la population ayant atteint chaque niveau désagrégé d'éducation et le nombre d'années de scolarisation "y_i" correspondant à ce dernier :

$$MAS = \sum_{i=1}^n y_i p_i$$

avec :

- n : nombre de niveaux désagrégés d'éducation retenus. Dans notre cas, 22 niveaux sont considérés (tableau n°15).
- y_i : nombre d'années de scolarisation correspondant aux différents niveaux désagrégés d'éducation (0 pour les illettrés ; 1 année pour la première année du primaire ; 2 années pour la deuxième année du

primaire... ; 10 années pour le tronc commun du secondaire (ex première année du lycée dans l'ancienne structure du système éducatif marocain) ...20 pour la troisième année de doctorat).

- p_i : proportion de la population ayant atteint le niveau d'éducation i .

L'indice de Gini de l'éducation est aussi calculé en fonction du nombre d'années d'études (le plus haut niveau de scolarité atteint), et comptabilise l'ensemble des écarts absolus entre chaque couple d'individus rapporté sur le double de la moyenne d'années de scolarisation selon la formule suivante :

$$\text{Gini}^{\text{Edu}} = \frac{1}{2 \cdot \text{MAS}} \cdot \sum_{i=1}^n \sum_{j=1}^n p_i |y_i - y_j| p_j$$

où :

- n représente le nombre des niveaux d'éducation retenus. Dans notre cas, 22 niveaux sont considérés allant de l'analphabétisme au doctorat.
- p_i et p_j représentent les proportions de la population ayant atteint respectivement les niveaux d'éducation i et j .
- y_i et y_j sont les années de scolarisation correspondant respectivement aux niveaux d'éducation i et j .
- MAS est la moyenne d'années de scolarisation.

Par ailleurs, il y a lieu de souligner l'originalité de nos estimations (EVS, MAS, Gini de l'éducation et IDH) puisque c'est pour la première fois au Maroc que de telles données assez désagrégées sont mobilisées pour des échelons inférieurs du territoire (régions, provinces et communes).

Signalons que l'analyse dans cette première section de ce chapitre portera sur le niveau régional avec une décomposition par genre et par milieu aussi bien pour les inégalités éducatives que celles relatives au développement humain. Elle concerne précisément 14 régions marocaines au lieu des 16 régions du découpage administratif puisque certaines sources de données - notamment dans l'ENDPR - regroupent les trois régions du sud en une seule région, appelée « région du Sud ».

3.2.2. Les inégalités régionales d'éducation au Maroc 1982-2010

L'éducation est étroitement liée à la dynamique de développement socio-économique des régions. En ce qui concerne la moyenne des années de scolarisation ou l'indice de Gini de l'éducation, les variations au niveau infra-national sont importantes. La décomposition de l'inégalité par milieu et genre permet d'évaluer leurs poids respectifs dans l'explication de l'inégalité éducative globale au sein des différentes régions.

3.2.2.1. Profil régional des inégalités d'éducation : un lent processus de convergence

Le tableau n° 16 montre l'évolution des inégalités éducatives et de la moyenne d'années de scolarisation dans les 14 régions marocaines en 1982, 1994, 2004 et 2010. La dynamique des inégalités éducatives est différente d'une région à

une autre. En 1982, les régions du Sud de Souss Massa Draa, Guelmim Es-semara, Taza Al Hoceima Taounate, Marrakech Tensift Al-Haouz, Doukkala Abda et Tadla Azilal affichent les plus fortes inégalités éducatives avec un Gini de l'éducation supérieur à la moyenne nationale (de 0,85 à 0,89). Les régions de Tanger Tétouan, Chaouia Ouardigha et Gharb Chrarda Beni Hssen présentent des niveaux intermédiaires (entre 0,81 à 0,83), en étant légèrement au-dessus de la moyenne nationale. Les inégalités les plus faibles sont observées au Grand Casablanca, Rabat Salé Zemmour Zear, Fès Boulemane, et dans une moindre mesure à Meknès Tafilalet et L'Oriental avec un indice de Gini, qui demeure tout de même relativement élevé et qui se situe entre 0,61 et 0,79.

Tableau 16: Indice de Gini et moyenne d'années de scolarisation par région

Région	Gini				MAS			
	1982	1994	2004	2010	1982	1994	2004	2010
Régions du Sud	0,85	0,69	0,59	0,52	1,55	3,12	4,46	5,49
Souss Massa Draa	0,87	0,78	0,68	0,63	1,17	2,16	3,29	3,85
Gharb Chrarda Beni Hssen	0,81	0,73	0,67	0,62	1,96	2,80	3,49	4,03
Chaouia Ouardigha	0,83	0,72	0,65	0,60	1,62	2,87	3,68	4,20
Marrakech Tensift Al-Haouz	0,87	0,80	0,71	0,65	1,39	2,13	3,07	3,85
L'Oriental	0,79	0,70	0,63	0,57	2,14	3,12	3,92	4,66
Grand Casablanca	0,61	0,53	0,47	0,43	4,05	5,23	6,21	6,93
Rabat Salé Zemmour Zear	0,69	0,60	0,54	0,49	3,53	4,64	5,63	6,19
Doukkala Abda	0,85	0,78	0,70	0,65	1,73	2,34	3,14	3,77
Tadla Azilal	0,86	0,77	0,71	0,65	1,41	2,41	3,06	3,65
Meknès Tafilalet	0,79	0,70	0,62	0,57	2,07	3,22	4,17	4,79
Fès Boulemane	0,76	0,68	0,61	0,56	2,53	3,46	4,43	5,17
Taza Al Hoceima Taounate	0,89	0,81	0,73	0,68	1,06	1,96	2,75	3,21
Tanger Tétouan	0,81	0,73	0,65	0,58	2,03	2,71	3,70	4,51
Maroc	0,79	0,71	0,63	0,58	2,13	3,13	4,05	4,72

Source : calculs de l'auteur à partir des données du recensement général de la population de 1982, 1994 et 2004 et de l'ENDPR 2009-2010.

Par ailleurs, les régions qui présentent une moyenne d'années de scolarisation relativement élevée ont une répartition relativement plus équitable de l'éducation mais le coefficient de Gini de l'éducation reste supérieur à 50% dans la plupart des cas. Les régions du Grand Casablanca, Rabat Salé Zemmour Zear, Fès Boulemane et Meknès Tafilalet présentent des niveaux de scolarité supérieurs à la moyenne nationale.

En dynamique, les inégalités ont légèrement baissé dans l'ensemble des régions entre 1982 et 2010. Cependant, les évolutions demeurent contrastées. Elles décroissent par exemple de manière substantielle dans les régions du Sud. Ces régions partent des niveaux d'inégalités les plus élevés en 1982 pour atteindre en 2010 les niveaux des plus faibles.

L'analyse de l'évolution des niveaux de scolarité montre que l'écart entre les régions (la différence entre la valeur maximale et minimale) évolue de façon croissante depuis les années 1980. Il passe de 2,99 années de scolarisation à 3,72 en 2010. Dans un contexte de fortes inégalités au niveau national, certaines régions connaissent une baisse plus prononcée des inégalités que d'autres en parallèle à une hausse de la moyenne des années de scolarisation.

Une analyse démographique par niveaux d'instruction maximum atteint nous permet de mettre en évidence les caractéristiques du système éducatif régional : la répartition de la population par niveau d'instruction montre en effet que les niveaux d'inégalité relativement faibles dans les régions du Grand Casablanca, Rabat Salé Zemmour Zear, les régions du Sud, Meknès Tafilalet, Fès Boulemane, sont à associer à leurs niveaux de scolarité plus élevés. Il s'agit des premières régions dont les taux d'alphabétisation ont fortement baissé et une large frange de la population a atteint des niveaux supérieurs.

Le développement de la scolarisation au cours des dernières années a permis d'augmenter la proportion de la population scolarisée sans pour autant réduire les écarts entre régions.

Les disparités dans la répartition de l'éducation par niveau scolaire, entre les différentes régions sont de plus en plus visibles aux cycles supérieurs. Six régions présentent des proportions d'illettrés inférieures à la moyenne nationale (Tableau n° 17). Bien que les niveaux de scolarité atteints au niveau primaire sont, dans l'ensemble, similaires entre régions, l'écart de scolarisation augmente à mesure que l'on passe aux niveaux supérieurs (collégial, secondaire et supérieur).

En outre, les régions où la moyenne d'années de scolarisation est importante sont celles où les proportions de la population ayant atteint des niveaux d'enseignement supérieur sont élevés et également où les proportions d'analphabètes sont faibles (elle est inférieure à la moyenne nationale notamment dans les régions du Grand Casablanca, Rabat Salé Zemmour Zear, Meknès Tafilalet, Fès Boulemane et dans celles du Sud). Ces dernières réalisent de bonnes performances compte tenu des proportions importantes de la population ayant un niveau collégial ou secondaire. Enfin, la région de Tanger Tétouan présente une moyenne d'année de scolarisation de 4,51 années en 2010 en raison de la faible proportion de la population ayant atteint le secondaire (7%) par rapport aux autres régions.

Tableau 17: Répartition de la population âgée de 15 ans et plus par niveau d'éducation atteint par région

Région	Néant				Préscolaire				Primaire				Collégial				Secondaire				Supérieur			
	1982	1994	2004	2010	1982	1994	2004	2010	1982	1994	2004	2010	1982	1994	2004	2010	1982	1994	2004	2010	1982	1994	2004	2010
Régions du Sud	74,3	54,5	42,1	35,8	5,8	6,0	4,2	2,2	9,1	17,0	19,4	18,7	5,7	11,8	16,9	17,4	3,3	7,5	12,4	19,7	1,8	3,3	5,0	6,2
Souss Massa Draa	78,6	66,4	53,2	46,8	5,6	5,2	3,6	3,6	7,9	13,8	20,4	22,8	4,3	7,5	11,6	12,6	2,4	4,6	7,4	9,9	1,2	2,5	3,8	4,1
Gharb Chrarda Beni Hssen	70,5	60,7	52,4	45,2	4,7	3,9	3,4	3,6	11,0	15,7	19,2	22,2	6,6	9,6	12,2	13,8	4,3	6,6	8,5	11,6	2,9	3,4	4,3	3,7
Chaouia Ouardigha	75,1	61,4	51,4	44,0	3,4	2,8	2,2	2,9	9,4	14,9	19,4	22,6	6,9	10,7	13,7	14,5	4,1	6,7	9,3	11,3	1,2	3,5	4,1	4,6
Marrakech Tensift Al-Haouz	78,5	69,1	57,4	48,8	4,7	4,4	3,6	4,2	6,5	11,4	17,5	19,8	5,1	7,2	9,9	11,9	3,0	5,0	7,3	10,5	2,2	2,9	4,2	4,9
l'Oriental	68,9	57,3	47,4	40,5	2,9	2,8	2,4	1,4	12,9	18,0	22,6	25,2	7,8	10,8	13,4	15,0	4,6	7,0	9,0	12,1	2,9	4,0	5,2	5,8
Grand Casablanca	44,9	34,5	27,8	23,5	4,1	3,4	2,1	1,8	19,8	23,4	23,3	22,3	16,5	17,4	19,4	17,5	10,1	14,1	17,2	24,2	4,5	7,2	10,2	10,7
Rabat Salé Zemmour Zear	55,2	43,9	36,1	31,5	3,8	3,5	2,2	1,7	14,2	18,5	19,4	19,5	12,2	14,0	16,9	17,1	8,3	11,7	14,8	20,4	6,4	8,3	10,7	9,8
Doukkala Abda	76,1	67,6	57,6	51,2	3,9	3,2	2,8	2,4	7,5	12,3	17,0	18,9	5,8	8,3	11,0	12,7	3,2	5,6	8,0	11,5	3,4	3,0	3,7	3,3
Tadla Azilal	78,6	67,8	58,7	51,4	3,6	2,5	2,1	2,4	7,2	11,8	16,8	19,5	5,4	8,8	11,7	13,3	3,5	5,9	7,2	9,6	1,6	3,2	3,4	3,9
Meknès Tafilalet	68,9	56,9	46,9	42,1	3,5	2,7	1,7	1,2	12,6	17,3	20,8	20,4	8,1	11,2	14,8	15,9	4,9	7,4	10,1	13,7	2,0	4,5	5,7	6,6
Fès Boulemane	64,5	55,0	45,1	39,5	4,0	2,9	2,0	1,3	12,2	17,0	20,3	20,8	9,4	11,6	14,7	15,8	6,2	8,0	10,8	13,9	3,6	5,5	7,1	8,6
Taza Al Hoceima Taounate	81,7	71,0	60,6	54,2	4,4	3,9	3,4	3,4	6,4	11,5	16,7	20,2	4,1	7,0	10,2	11,8	2,2	4,2	5,9	7,5	1,2	2,4	3,1	2,8
Tanger Tétouan	66,2	56,8	46,3	36,8	9,2	8,9	7,6	8,2	10,2	15,5	19,9	22,1	6,6	9,3	12,7	16,0	4,1	6,0	8,2	11,1	3,6	3,5	5,3	5,7
Maroc	68,7	57,5	47,6	41,1	4,5	4,0	3,1	3,0	11,0	16,0	19,8	21,3	8,0	10,7	13,7	14,7	4,9	7,5	10,1	13,8	2,9	4,3	5,8	6,1

Source : Calculs de l'auteur à partir des données des recensements de la population du HCP, 1982, 1994, 2004 et de l'ENDPR 2009-2010

Entre 1982 et 2010, la durée moyenne de scolarisation au niveau national a été multipliée par 2. Cette augmentation est inégalement répartie entre les régions : 4 années de scolarisation de plus pour les régions du Sud, moins de 2,5 années pour les régions Doukkala Abda, Gharb Chrarda Beni Hssen, Taza Al Hoceima Taounate, Tadla Azilal, Marrakech Tensift Al-Haouz et Tanger Tétouan (Tableau n° 17).

Quatre groupes peuvent être distingués selon la valeur de la moyenne d'années de scolarisation en 2010. Le premier groupe est constitué du Grand Casablanca et de Rabat Salé Zemmour Zear qui présentent une durée moyenne de scolarisation supérieure à 6 années. Le deuxième groupe comprend les régions du Sud, Fès Boulemane et de Meknès Tafilalet avec une moyenne supérieure à celle enregistrée au niveau national (4,72 années) et inférieure à 6 années de scolarisation. Le troisième groupe réunit les régions de l'Oriental, Tanger Tétouan, Chaouia Ouardigha et Gharb Chrarda Beni Hssen, et se distingue par une moyenne inférieure à la moyenne nationale mais qui demeure supérieure à 4 années de scolarisation. Le quatrième groupe, est composé des régions de Souss Massa Draa, Marrakech Tensift Al-Haouz, Doukkala Abda, Tadla Azilal, et Taza Al Hoceima Taounate. Il présente à la fois un niveau moyen d'éducation faible (inférieur à 4 années) et un indice d'inégalité de l'éducation assez élevé (supérieur à 0,63).

La région de Taza Al Hoceima Taounate présente le niveau de scolarité le plus faible, et ce depuis 1982. La région du Grand Casablanca conserve la première position avec un niveau moyen de scolarisation élevé. L'écart entre ces deux régions ne cesse de se creuser au fil du temps (il est de 2,99 années en 1982 ; de 3,27 années en 1994, de 3,45 années en 2004 et de 3,72 années en 2010). Les différences absolues entre les régions ne sont pas restées les mêmes entre 1982 et 2010. Mise à part les régions du Sud qui, partant des niveaux de scolarisation faibles, ont connu une croissance plus forte en termes de durée de scolarisation. Les autres régions qui au départ avaient une faible dotation éducative n'ont pas connu des rythmes de croissance soutenus. Ce schéma d'évolutions contrastées entre les régions montre toutefois un processus de divergence.

En 1982, le taux d'analphabétisme dans les régions marocaines atteignait 82% (enregistré pour la région de Taza Al Hoceima Taounate). Par contre, il a fortement baissé en 2010 et se situe entre 23% et 54%. Notons que les disparités entre les régions en termes de proportions de la population ayant atteint le niveau primaire, plus visible en 1982, ont largement baissées en 2010. Par contre, les régions divergent en ce qui concerne la proportion de la population qui atteint le niveau secondaire. L'expansion de l'enseignement secondaire a concerné principalement les régions du sud, du Grand Casablanca, Rabat Salé Zemmour Zear, Meknès Tafilalet et Fès Boulemane.

3.2.2.2. Des inégalités de genre et de milieu plus prononcées dans les régions les moins développées

Le tableau n° 18 donne l'évolution de l'indice de Gini régional par genre. De manière générale, les femmes sont confrontées à des niveaux d'inégalités éducatives plus élevées que les hommes particulièrement dans les régions de Taza Al Hoceima Taounate, Souss Massa Draa, Marrakech Tensift Al-Haouz, Doukkala Abda et Tadla Azilal). Les inégalités éducatives régionales de genre ont légèrement

reculé sur la période 1982-2010 mais demeurent relativement persistantes. Les régions du Grand Casablanca, Rabat Salé Zemmour Zear et les régions du Sud affichent des niveaux d'inégalités faibles auprès des femmes en 2010. Entre 1982 et 2010, ce sont les régions du sud qui ont accompli des avancées notables en matière de réduction des inégalités entre les femmes.

Tableau 18 : Indice de Gini régional de l'éducation par genre

Région	1982		1994		2004		2010	
	h	f	h	f	h	f	h	f
Régions du Sud	0,75	0,93	0,58	0,80	0,50	0,67	0,42	0,61
Souss Massa Draa	0,78	0,95	0,67	0,89	0,56	0,78	0,52	0,73
Gharb Chrarda Beni Hssen	0,74	0,88	0,64	0,81	0,58	0,76	0,52	0,71
Chaouia Ouardigha	0,76	0,90	0,64	0,81	0,56	0,73	0,51	0,69
Marrakech Tensift Al-Haouz	0,82	0,91	0,73	0,86	0,64	0,78	0,57	0,72
L'Oriental	0,70	0,87	0,61	0,79	0,54	0,71	0,48	0,65
Grand Casablanca	0,53	0,68	0,46	0,59	0,41	0,53	0,36	0,49
Rabat Salé Zemmour Zear	0,62	0,76	0,53	0,68	0,46	0,61	0,41	0,56
Doukkala Abda	0,79	0,90	0,71	0,84	0,63	0,78	0,56	0,73
Tadla Azilal	0,79	0,93	0,69	0,85	0,61	0,79	0,55	0,75
Meknès Tafilalet	0,71	0,87	0,60	0,79	0,52	0,70	0,48	0,66
Fès Boulemane	0,68	0,83	0,60	0,77	0,52	0,69	0,47	0,64
Taza Al Hoceima Taounate	0,81	0,95	0,70	0,90	0,62	0,83	0,56	0,79
Tanger Tétouan	0,74	0,87	0,66	0,80	0,58	0,72	0,51	0,65
Maroc	0,72	0,86	0,62	0,79	0,55	0,71	0,49	0,66

h : hommes ; f : femmes.

Source : calculs des auteurs à partir des données des recensements de la population du HCP, 1982, 1994, 2004 et de l'ENDPR 2009-2010.

Certaines régions affichent des différences non négligeables entre sexe, à l'avantage des hommes, notamment au Gharb Chrarda Beni Hssen, Chaouia Ouardigha, Marrakech Tensift Al-Haouz, Tadla Azilal, Meknès Tafilalet et l'Oriental. Plus précisément, l'écart entre les proportions de la population selon les niveaux d'alphabétisation des hommes et des femmes ne s'est pas résorbé dans le temps, ce qui s'observe directement dans les taux d'alphabétisation. L'évolution de la répartition de la population par niveau d'éducation atteint depuis le recensement de 1982 montre que dans l'ensemble des régions, l'alphabétisation profite d'abord aux hommes. Comparativement aux autres régions, les régions de Taza Al Hoceima Taounat, Souss Massa Draa, Marrakech Tensift Al-Haouz, Doukkala Abda et Tadla Azilal se classent parmi les régions où les inégalités liées de genre sont les plus prononcées. Les régions comme le Grand Casablanca, Rabat Salé Zemmour Zear et les régions du Sud affichent de bonnes performances avec notamment un rapprochement des niveaux d'inégalités des femmes et des hommes et une certaine parité au primaire et dans une moindre mesure au secondaire.

L'analyse des inégalités d'éducation selon le milieu révèle que les inégalités restent nettement plus prononcées dans le milieu rural que dans le milieu urbain (tableau n° 19). De manière générale, on note une certaine baisse dans le milieu urbain et dans le milieu rural dans l'ensemble des régions entre 1982 et 2010. En 2010, six régions sur quatorze présentent des niveaux d'inégalités dans le milieu rural encore plus élevées que la moyenne nationale. Lesquelles sont le Gharb Chrarda Beni Hssen, Marrakech Tensift Al-Haouz, Doukkala Abda, Tadla Azilal,

Taza Al Hoceima Taounate et Tanger Tétouan. De manière particulière, l'indice de Gini a plus fortement baissé dans le milieu urbain dans les régions du sud et dans le rural dans Souss Massa Draa, Grand Casablanca, Meknès Tafilalet et les régions du sud.

Tableau 19 : Indice de Gini régional de l'éducation par milieu rural/urbain

Région	1982		1994		2004		2010	
	u	r	u	r	u	r	u	r
Régions du Sud	0,79	0,90	0,63	0,84	0,54	0,72	0,47	0,64
Souss Massa Draa	0,69	0,92	0,63	0,86	0,55	0,76	0,52	0,71
Gharb Chrarda Beni Hssen	0,63	0,90	0,56	0,85	0,52	0,78	0,48	0,71
Chaouia Ouardigha	0,67	0,91	0,57	0,83	0,51	0,76	0,48	0,70
Marrakech Tensift Al-Haouz	0,67	0,95	0,61	0,90	0,54	0,82	0,49	0,76
L'Oriental	0,65	0,90	0,60	0,83	0,55	0,76	0,50	0,68
Grand Casablanca	0,59	0,81	0,52	0,70	0,46	0,61	0,42	0,54
Rabat Salé Zemmour Zear	0,61	0,89	0,55	0,83	0,49	0,75	0,45	0,66
Doukkala Abda	0,65	0,93	0,57	0,89	0,50	0,82	0,45	0,76
Tadla Azilal	0,70	0,92	0,63	0,85	0,58	0,78	0,55	0,71
Meknès Tafilalet	0,65	0,89	0,58	0,82	0,52	0,73	0,48	0,70
Fès Boulemane	0,67	0,90	0,62	0,83	0,55	0,75	0,50	0,71
Taza Al Hoceima Taounate	0,65	0,92	0,61	0,86	0,55	0,79	0,53	0,72
Tanger Tétouan	0,68	0,92	0,62	0,89	0,55	0,81	0,49	0,71
Maroc	0,64	0,92	0,58	0,86	0,52	0,78	0,47	0,71

u : urbain ; r : rural.

Source : calculs des auteurs à partir des données des recensements de la population du HCP, 1982, 1994, 2004 et de l'ENDPR 2009-2010.

Si l'on regarde de façon plus spécifique le niveau des inégalités par couple genre/milieu, il en ressort que le groupe le plus défavorisé est constitué des femmes rurales particulièrement localisées dans les régions de Doukkala Abda, Taza Al Hoceima Taounate et Marrakech Tensift Al-Haouz. Ce groupe présente des chances d'accès aux différents niveaux d'études extrêmement faibles par rapport à celles du groupe le plus favorisé constitué des hommes vivant en milieu urbain notamment dans les régions de Chaouia Ouardigha, Grand Casablanca, Rabat Salé Zemmour Zear, Doukkala Abda et Meknès Tafilalet (tableau n° 20).

En 2010, la différence atteint 6 années de scolarisation (7,28 pour les hommes urbains contre 1,18 pour les femmes rurales pour la région de Doukkala Abda). Notons qu'en 2010, la région du Grand Casablanca présente l'indice de Gini de l'éducation des femmes rurales le plus faible en raison notamment de la présence d'une forte proportion de la population ayant atteint le niveau secondaire (16% au collège et 10% au lycée) par rapport aux autres régions.

Ces quelques faits stylisés ont permis de mettre en évidence un certain nombre de caractéristiques du système éducatif au niveau régional. Les inégalités éducatives des différentes régions sont généralement plus importantes pour les femmes que pour les hommes, dans le milieu rural que dans le milieu urbain.

Tableau 20 : Indice de Gini régional selon le genre et le milieu de résidence

Région	1982				1994				2004				2010			
	h/u	h/r	f/u	f/r												
Régions du Sud	0,72	0,78	0,88	0,99	0,54	0,71	0,73	0,93	0,46	0,61	0,62	0,81	0,39	0,51	0,55	0,73
Souss Massa Draa	0,58	0,83	0,81	0,99	0,53	0,75	0,73	0,96	0,46	0,63	0,65	0,87	0,43	0,59	0,60	0,80
Gharb Chrarda Beni Hssen	0,54	0,84	0,72	0,97	0,48	0,76	0,64	0,94	0,44	0,68	0,60	0,88	0,40	0,61	0,57	0,82
Chaouia Ouardigha	0,55	0,85	0,77	0,97	0,47	0,74	0,66	0,92	0,42	0,66	0,59	0,85	0,38	0,59	0,57	0,80
Marrakech Tensift Al-Haouz	0,59	0,90	0,75	0,98	0,53	0,84	0,69	0,96	0,46	0,74	0,61	0,90	0,41	0,67	0,56	0,83
L'Oriental	0,55	0,81	0,74	0,97	0,50	0,74	0,68	0,92	0,46	0,66	0,62	0,85	0,42	0,58	0,58	0,78
Grand Casablanca	0,52	0,72	0,66	0,91	0,45	0,61	0,58	0,81	0,39	0,52	0,52	0,70	0,35	0,46	0,47	0,63
Rabat Salé Zemmour Zear	0,54	0,83	0,69	0,95	0,47	0,74	0,62	0,92	0,42	0,65	0,55	0,85	0,38	0,55	0,52	0,76
Doukkala Abda	0,56	0,89	0,74	0,96	0,49	0,82	0,65	0,96	0,43	0,73	0,58	0,91	0,37	0,67	0,53	0,86
Tadla Azilal	0,60	0,86	0,81	0,98	0,53	0,77	0,72	0,93	0,48	0,69	0,66	0,87	0,45	0,61	0,65	0,81
Meknès Tafilalet	0,55	0,81	0,74	0,97	0,49	0,72	0,67	0,91	0,43	0,63	0,61	0,83	0,39	0,60	0,56	0,79
Fès Boulemane	0,59	0,82	0,75	0,97	0,53	0,73	0,70	0,93	0,47	0,64	0,62	0,86	0,42	0,59	0,58	0,82
Taza Al Hoceima Taounate	0,54	0,85	0,75	0,98	0,49	0,76	0,71	0,95	0,44	0,67	0,64	0,89	0,43	0,60	0,62	0,84
Tanger Tétouan	0,60	0,85	0,76	0,98	0,54	0,81	0,70	0,97	0,48	0,73	0,61	0,89	0,43	0,63	0,55	0,79
Maroc	0,56	0,85	0,73	0,98	0,49	0,77	0,66	0,94	0,44	0,68	0,59	0,87	0,40	0,61	0,55	0,81

h/u : hommes urbains ; h/r : hommes ruraux ; f/u : femmes urbaines ; f/r : femmes rurales.

Source : calculs des auteurs à partir des données des recensements de la population du HCP, 1982, 1994, 2004 et de l'ENDPR 2009-2010.

3.2.3. Les disparités régionales de développement au Maroc

3.2.3.1. Le niveau de développement humain des régions marocaines

Le tableau n° 21 donne les valeurs de l'IDH et de ses composantes. De fortes disparités entre les régions marocaines sont mises en évidence : l'IDH varie de 0,473 pour la région de Taza Al Hoceima Taounate à 0,709 pour la région du Grand Casablanca. L'écart relatif de développement humain entre ces deux régions atteint 33%. Il apparaît cependant plus prononcé dans l'éducation (51%) et dans une moindre mesure pour la richesse monétaire (27%) et pour la santé (17%). Les valeurs d'IDH les plus élevées sont observées dans la région Rabat Salé Zemmour Zaer et dans celles du Sud. Les régions ayant un IDH faible sont Tadla Azilal, Gharb Chrarda Beni Hssen, Doukkala Abda, Marrakech Tensift Al-Haouz et Souss Massa Draa. Meknès Tafilalet, Tanger Tétouan, Chaouia Ouardigha et l'Oriental ont un IDH inférieur à la moyenne nationale (soit 0,569, 0,573, 0,585 et 0,579 respectivement). La région de Fès Boulemane a un IDH légèrement supérieur à la moyenne nationale (0,593).

Les régions Rabat Salé Zemmour Zaer, Sud et Fès Boulemane présentent un niveau de développement moyen (dont les niveaux d'IDH oscillent autour de 0,631). Les régions Chaouia Ouardigha, l'Oriental, Tanger Tétouan, Meknès Tafilalet, Souss Massa Draa, Marrakech Tensift Al-Haouz, Doukkala Abda et Gharb Chrarda Beni Hssen seraient situées au rang des pays à niveau de développement humain moyen. Enfin, Tadla Azilal et Taza Al Hoceima Taounate ont un niveau de développement humain faible (dont l'IDH est compris entre 0,525 et 0,295).

L'analyse par composante montre que les régions du Grand Casablanca, de Rabat Salé Zemmour Zear, de Fès Boulemane, du Sud et de Meknes Tafilalet occupent les cinq premiers rangs en ce qui concerne l'indice de l'éducation. Ceci est valable aussi lorsque l'on considère les sous composantes de l'indice éducation, à savoir la moyenne des années de scolarisation des 25 ans et plus et l'espérance de vie scolaire. De manière générale, ces régions présentent en moyenne des niveaux de scolarité élevés, en raison notamment des taux d'analphabétisme relativement bas, de la présence d'une forte proportion de la population ayant atteint des niveaux supérieurs et d'un faible taux de déperdition.

Les disparités entre les régions apparaissent marquées lorsqu'on considère le niveau de vie. En termes d'indice de revenu, les cinq premières régions, qui présentent des niveaux de revenu régional brut par tête supérieur à la moyenne nationale (compris entre 4381 et 6648 en \$ constant PPA 2005), se caractérisent par une prédominance du secteur secondaire et/ou tertiaire. En effet, les régions du Grand Casablanca et de Rabat Salé Zemmour Zaer ont un revenu plus élevé par rapport aux autres régions marocaines (6648 et 6515 en PPA \$ constant 2005 respectivement). Ces deux régions se caractérisent par une prépondérance des activités industrielles et de services (98,4% et 92,1% respectivement). Les régions du Sud sont plutôt à dominance tertiaire (soit 63% du PIB de la région), les situant en deuxième position après Rabat-Salé Zemmour Zaer dont le secteur tertiaire représente 72% du PIB de la région.

Tableau 21: Valeurs de l'IDH et de ses composantes, régions marocaines (2010)

Région	Les indicateurs				Les indices						
	EVN	MAS	EVS	RRB	ISanté	IMAS	IEVS	IEducation	IRevenu	IDH	IDH _{nm}
	<i>En années</i>			<i>En PPA \$ constant 2005</i>							
Régions du Sud	76,32	4,38	10,97	5992	0,89	0,33	0,61	0,46	0,60	0,628	0,640
Souss Massa Draa	77,01	2,79	10,66	2975	0,90	0,21	0,59	0,36	0,50	0,547	0,571
Gharb Chrarda Beni Hssen	75,22	3,30	9,77	2623	0,87	0,25	0,54	0,38	0,48	0,541	0,573
Chaouia Ouardigha	73,24	3,35	10,64	5825	0,84	0,25	0,59	0,40	0,60	0,585	0,577
Marrakech Tensift Al-Haouz	73,54	3,01	9,99	3616	0,84	0,23	0,56	0,37	0,53	0,546	0,555
L'Oriental	75,95	3,82	10,56	3435	0,88	0,29	0,59	0,42	0,52	0,579	0,610
Grand Casablanca	79,25	6,24	13,72	6648	0,93	0,47	0,76	0,62	0,62	0,709	0,758
Rabat Salé Zemmour Zear	76,57	5,40	12,09	6515	0,89	0,41	0,67	0,54	0,62	0,666	0,692
Doukkala Abda	70,87	3,09	9,60	4381	0,80	0,23	0,53	0,36	0,56	0,545	0,539
Tadla Azilal	71,17	2,61	10,36	2529	0,80	0,20	0,58	0,35	0,48	0,511	0,528
Meknès Tafilalet	71,07	3,91	11,33	3338	0,80	0,29	0,63	0,44	0,52	0,569	0,597
Fès Boulemane	73,17	4,34	11,93	3385	0,84	0,33	0,66	0,48	0,52	0,593	0,633
Taza Al Hoceima Taounate	69,21	2,32	8,84	2171	0,77	0,17	0,49	0,30	0,45	0,473	0,483
Tanger Tétouan	73,57	3,70	10,15	4044	0,84	0,28	0,56	0,41	0,55	0,573	0,587
Maroc	74,80	3,90	10,83	4189	0,86	0,29	0,60	0,43	0,55	0,590	0,611
HDR Maroc	71,87	4,37	10,36	4127	0,82	0,33	0,58	0,45	0,55	0,586	0,605
Ecart relatif Maximum* (%)	-	-	-	-	17	63	36	51	27	33	36

EVN : Espérance de vie à la naissance; MAS: Moyenne d'années de scolarisation; EVS : Espérance de vie scolaire; RRB : Revenu régional Brut ; IDH : Indice de développement humain ; IDH_{nm} : Indice de développement humain non monétaire.

* L'écart relatif maximum dans un indice donné est égal à la différence entre l'indice du Grand Casablanca et celui de Taza Al Hoceima Taounate rapportée à l'indice du Grand Casablanca.

Source : Calculs de l'auteur

La part du secteur secondaire est forte dans les régions de Chouia-Ouardigha et de Doukala-Abda (53% et 41,5% du PIB respectivement). Par contre, Taza Al Hoceïma Taounate, Tadla Azilal et Gharb Chrarda Béni Hssen affichent des niveaux de revenu régional brut par tête très faibles (entre 2171 et 2623 en PPA \$ constant 2005) et des parts relativement élevés du secteur primaire dans le PIB (soit 33,3%, 35,4% et 31,4% respectivement).

Les disparités en matière de santé sont relativement moins importantes que les deux autres dimensions de l'IDH (l'indice de la santé se situe entre 0,77 pour la région de Taza Al Houceïma Taounate et 0,93 pour la région du Grand Casablanca). Toutefois, l'espérance de vie à la naissance de la région du Grand Casablanca est supérieure de 10 ans à celle de la région Taza Al Houceïma Taounate (79,2 et 69,2 ans respectivement). Notons que l'espérance de vie à la naissance dans le Grand Casablanca est proche de la moyenne des pays à développement humain très élevé selon la classification du PNUD (80 ans), soit la 30^{ème} position mondiale dans la composante santé. Celle de Taza Al Houceïma Taounate est inférieure à la moyenne des pays à niveau de développement humain moyen (lui conférant la 114^{ème} position mondiale dans la composante santé) (voir tableau n° 22).

- ***Profil de développement social des régions marocaines***

Le classement des trois premières régions et des trois dernières régions ne change pas que l'on considère l'IDH ou l'indice de revenu. En revanche, les écarts de rang entre l'IDH et l'indice de revenu peuvent être significatifs pour les autres régions. En particulier, les régions de Doukala-Abda et de Marrakech Tensift Al-Haouz ont un rang meilleur pour le revenu régional brut par rapport à leur rang dans l'IDH (6 places et 3 places respectivement). Ces régions, malgré leur niveau de revenu, accusent un déficit dans les composantes sociales de l'IDH.

Au niveau régional, les situations des régions sont contrastées selon que l'on considère les valeurs de l'IDH non monétaire, de l'indice de santé ou de l'indice de l'éducation, par genre et par milieu (voir annexe n°1). Nous adoptons une approche qui consiste à classer les régions par rapport à la moyenne nationale, pour les indices de l'éducation et de santé ainsi que leurs déclinaisons par genre et milieu, afin de dégager des profils types regroupant les régions aux caractéristiques plus au moins similaires en matière de développement social.

En matière de développement social, quatre groupes de régions se distinguent. Le premier groupe est constitué des régions du Grand Casablanca, Rabat Salé Zemmour Zaer et du Sud qui présentent des indices relativement élevés aussi bien dans l'éducation que dans la santé par genre et milieu. Soulignons que, de manière générale, ce groupe connaît toutefois, en milieu rural, des indices de santé légèrement inférieurs à la moyenne nationale pour les deux sexes. Le deuxième groupe comprend les régions de Fès Boulemane et de Meknès Tafilalet. Il se caractérise par un IDH non monétaire et des indices de l'éducation assez proches de la moyenne nationale, aussi bien pour l'ensemble de la région qu'en milieu rural. En revanche, les indices de santé sont inférieurs à la moyenne nationale en milieu rural. En milieu urbain, ces deux régions ont des indices d'éducation et de santé inférieurs à la moyenne nationale.

Tableau 22: Rang de l'IDH et de ses composantes : classement national et international, régions marocaines (2010)

Région	Classement mondial							Classement national						
	I _{Santé}	I _{MAS}	I _{EVS}	I _{Education}	I _{Revenu}	IDH	IDH _{nm}	I _{Santé}	I _{MAS}	I _{EVS}	I _{Education}	I _{Revenu}	IDH	IDH _{nm}
Régions du Sud	47	156	133	146	100	122	127	4	3	5	4	3	3	3
Souss Massa Draa	40	176	141	166	135	137	140	2	12	6	11	11	9	10
Gharb Chrarda Beni Hssen	56	170	155	163	138	137	140	6	9	12	9	12	12	9
Chaouia Ouardigha	83	169	141	162	101	131	139	9	8	7	8	4	5	8
Marrakech Tensift Al-Haouz	79	175	155	166	124	137	144	8	11	11	10	7	10	11
L'Oriental	49	163	143	157	125	132	130	5	6	8	6	8	6	5
Grand Casablanca	30	127	64	118	99	94	91	1	1	1	1	1	1	1
Rabat Salé Zemmour Zear	44	140	108	133	100	110	119	3	2	2	2	2	2	2
Doukkala Abda	108	175	156	166	116	137	145	13	10	13	12	5	11	12
Tadla Azilal	107	178	148	171	138	147	148	11	13	9	13	13	13	13
Meknès Tafilalet	107	161	130	152	126	134	135	12	5	4	5	10	8	6
Fès Boulemane	87	157	113	144	125	130	127	10	4	3	3	9	4	4
Taza Al Hoceima Taounate	114	179	164	178	142	152	158	14	14	14	14	14	14	14
Tanger Tétouan	79	165	151	158	120	134	137	7	7	10	7	6	7	7
Maroc	59	162	137	155	116	130	130							
HDR Maroc	105	156	146	150	117	130	131							

MAS: Moyenne d'années de scolarisation; EVS : Espérance de vie scolaire; IDH : Indice de développement humain ; IDHnm : Indice de développement humain non monétaire.

Source : Calculs des auteurs. Les indices (I_{santé}, I_{MAS}, I_{EVS}, I_{Education}, I_{Revenu} et IDH_{nm}) des 187 pays du monde, normalisés selon les *goalposts* 2013, ne sont pas disponibles pour l'année 2010 dans le RDH ,2013. Nous les avons calculé nous-même afin d'identifier le classement des régions marocaines

Le troisième groupe, qui réunit les régions de l'Oriental, Chaouia Ouardigha, Gharb Chrarda Beni Hssen et Souss Massa Draa, se distingue par un IDH non monétaire supérieur à la moyenne nationale uniquement en milieu rural. Contrairement aux indices de santé qui enregistrent des niveaux relativement élevés particulièrement en milieu rural, les indices d'éducation de ce groupe sont souvent inférieurs à la moyenne nationale aussi bien en milieu rural qu'en milieu urbain. Le quatrième groupe, est composé des régions de Tanger Tétouan, Marrakech Tensift Al-Haouz, Doukkala Abda, Tadla Azilal, et Taza Al Hoceima Taounate. Ces régions présentent à la fois des indices d'éducation et de santé souvent en dessous de la moyenne nationale, aussi bien en milieu urbain que rural.

3.2.3.2. Les écarts de développement entre régions

L'analyse en termes d'écart relatif peut sous-estimer l'écart de développement observé. De même, le recours au rang de classement, même s'il permet de mieux tenir compte de la concentration dans la distribution des valeurs de l'IDH, ne traduit pas pour autant l'écart de développement mesuré par l'IDH (Sudrie, 2012). En effet, si l'écart relatif séparant le Grand Casablanca de la région Rabat Salé Zemmour Zear est égal à celui séparant les régions du Sud de Fès Boulemane (soit un écart relatif de 6%), l'écart en termes de rang mondial entre les deux premières régions est égal au double de celui entre les deux dernières (soit un écart de 16 places [110-94] et 8 places [130-122] respectivement). De plus, en termes de comparaison internationale, une seule place sépare le Maroc (130^{ème}) du Guatemala (129^{ème}) et de l'Irak (131^{ème}). Toutefois, l'écart d'IDH qui sépare le Maroc et le Guatemala est de 0,008 alors qu'il est seulement de 0,001 entre le Maroc et l'Irak (soit huit fois moins que par rapport au Guatemala).

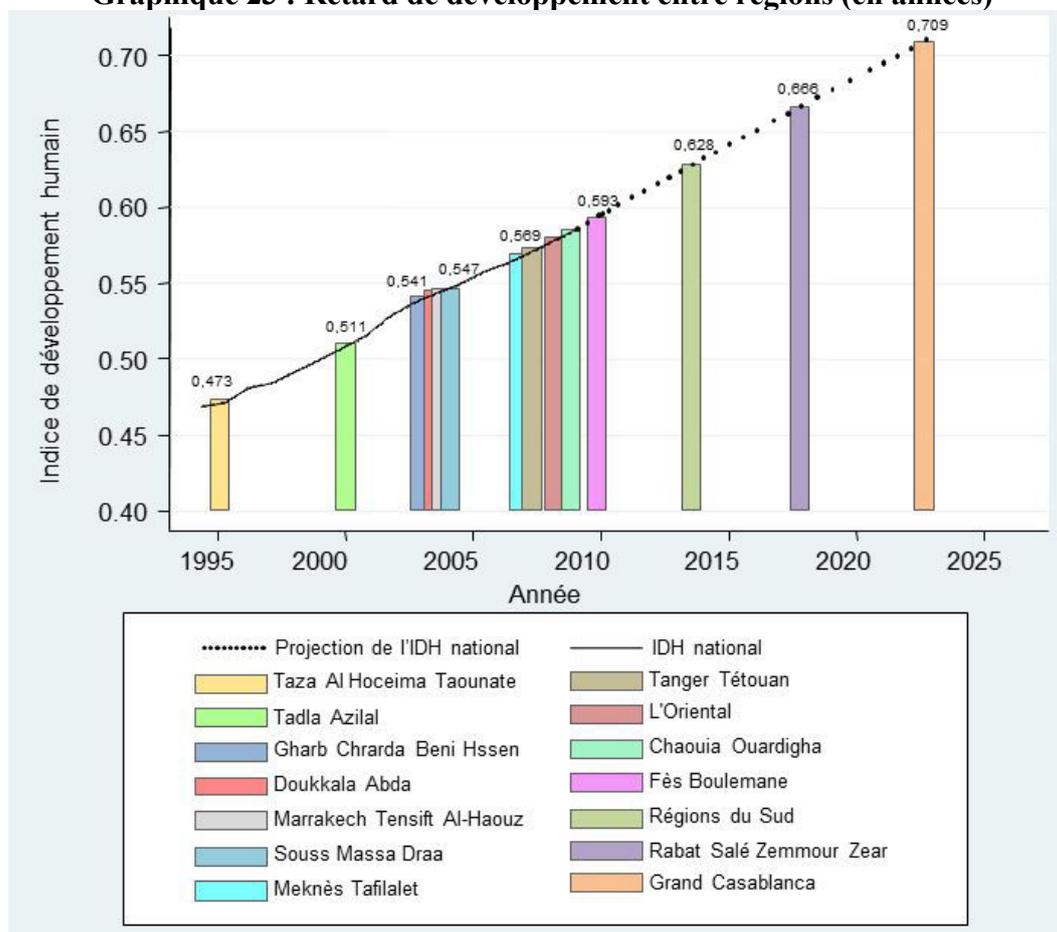
Pour mesurer les écarts de développement entre les régions, nous calculons le nombre d'années de retard en estimant l'année à laquelle le Maroc avait atteint un niveau de développement comparable à celui des régions ayant un IDH inférieur à la moyenne nationale. Pour les valeurs d'IDH national après l'année 2010, nous projetons les valeurs de chacune des composantes en extrapolant la tendance 2005-2010 (voir annexe n° 7). Ceci permet à la fois de situer chacune des régions dans la trajectoire d'évolution de l'IDH du Maroc et de calculer les années de retard entre les régions marocaines elles-mêmes.

Le graphique n° 21 reporte l'évolution de l'IDH national depuis 1995 et celui constaté en 2010 pour les régions marocaines. Ainsi, l'IDH de la région de Taza Al Hoceima Taounate et de Tadla Azilal se situe au même niveau que celui atteint par le Maroc en 1995 et en 2000. Soit un retard de quinze ans de développement pour la première et de dix ans pour la seconde. Les régions de Souss Massa Draa, Marrakech Tensift Al-Haouz, Doukkala Abda et Gharb Chrarda Beni Hssen, avec un IDH moyen de l'ordre de 0,545, accusent un retard compris entre six et sept années par rapport à l'IDH national. Les régions de Chaouia Ouardigha, l'Oriental, Tanger Tétouan et Meknès Tafilalet affichent, elles, un niveau de développement légèrement en deçà de la moyenne nationale avec un retard compris entre une et trois années.

D'après nos estimations, le taux d'accroissement annuel moyen de l'IDH entre 2005 et 2010 pour le Maroc est de l'ordre de 1,41%, légèrement supérieur à celui observé par les pays à développement humain moyen (1,38%). A cette vitesse,

le Maroc atteindra le niveau d'IDH du Grand Casablanca dans 13 années. Soit un retard de 28 années entre la région Taza Al Hoceima Taounate et du Grand Casablanca (voir graphique n° 23).

Graphique 23 : Retard de développement entre régions (en années)



Source : Auteur.

Le calcul de l'IDH national est basé sur les estimations (jusqu'en 2010) et les projections après 2010 des indicateurs (EVN, EVS, MAS et RNB) sur la période 1995-2025 (voir annexe n°2).

Les écarts de développement de l'IDH agrégé ainsi mis en évidence peuvent différer d'une composante à l'autre. Le tableau n° 11 donne, pour chaque composante de l'IDH, le nombre d'années de retard par rapport à la région du Grand Casablanca. C'est en termes d'années de scolarisation que les écarts des différentes régions par rapport au Grand Casablanca sont les plus significatifs. La région Taza Al Hoceima Taounate accuse un retard important dans toutes les composantes de l'IDH. Il est de 21 ans pour l'espérance de vie à la naissance, 41 ans pour la moyenne d'années de scolarisation, 17 ans pour l'espérance de vie scolaire et de 47 ans pour le revenu (Annexe n° 7).

Sur la base de données très désagrégées issues des recensements de la population marocaine (1982, 1994 et 2004), de l'ENDPR 2009-2010, et des recensements scolaires et universitaires, notre but est ici d'évaluer au niveau régional le niveau de développement humain et des inégalités éducatives en estimant l'indice de Gini de l'éducation et l'IDH selon la nouvelle méthodologie du PNUD (2010).

Dans cette section consacrée au niveau régional, on a montré l'existence de fortes inégalités éducatives et d'énormes disparités de développement humain. Ces dernières s'accroissent en considérant, en sus de la dimension régionale, une déclinaison par milieu et par genre.

Tableau 23: Retard en développement humain par rapport à la région du Grand Casablanca (en année)

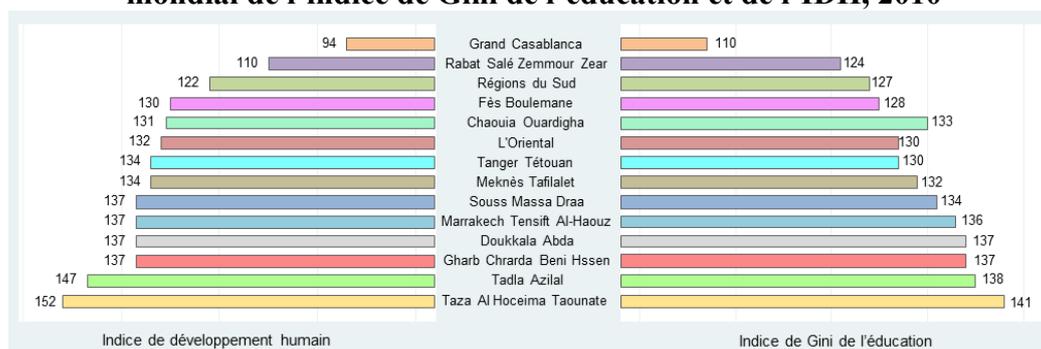
Région	EVN	MAS	EVS	RNB	IDH
Régions du Sud	5	19	8	3	10
Souss Massa Draa	4	36	9	23	19
Gharb Chrarda Beni Hssen	7	30	15	28	20
Chaouia Ouardigha	11	30	9	4	14
Marrakech Tensift Al-Haouz	11	33	14	18	19
L'Oriental	6	25	10	19	15
Rabat Salé Zemmour Zear	5	9	5	1	5
Doukkala Abda	16	32	16	12	20
Tadla Azilal	15	37	10	35	23
Meknès Tafilalet	16	24	7	20	16
Fès Boulemane	11	19	6	20	13
Taza Al Hoceima Taounate	21	41	17	47	28
Tanger Tétouan	10	26	12	15	16
Maroc	8	24	8	13	13

Source : Auteur.

Ainsi, certaines régions marocaines accusent un retard important dans les différentes composantes de développement humain par rapport à la région du Grand Casablanca. C'est le cas notamment de la région Taza Al Hoceima Taounate dont le retard atteint 21 ans pour l'espérance de vie à la naissance, 41 ans pour la moyenne d'années de scolarisation, 17 ans pour l'espérance de vie scolaire et de 47 ans pour le revenu. De manière générale, par rapport au Grand Casablanca, le retard est plus prononcé pour la majorité des régions en termes de moyenne d'années de scolarisation.

On a positionné les régions marocaines selon le classement de 2010 à partir de la Base mondiale des inégalités éducatives et à partir du classement international du PNUD (graphique n° 24). Une grande similitude peut être remarquée entre les deux classements. Ceci amène à se poser la question sur le lien qui peut exister entre les inégalités éducatives et le développement au niveau territorial.

Graphique 24 Rang des régions marocaines dans le classement mondial de l'indice de Gini de l'éducation et de l'IDH, 2010



Source : Auteur à partir de la Base mondiale des inégalités éducatives et de la base de l'IDH du PNUD.

3.3. Inégalités éducatives et développement humain au niveau communal : une approche par l'économétrie spatiale

Cette section est consacrée à l'effet des inégalités éducatives au niveau communal au Maroc, à partir d'une approche par l'économétrie spatiale.

Nous avons construit une base de données originale sur les 1534 communes marocaines en 2004, année pour laquelle des informations à ce niveau géographique peuvent être mobilisées. En plus de l'indice de Gini de l'éducation et de l'indice de développement humain. Rappelons que l'estimation de l'indice de Gini de l'éducation au niveau communal est basée sur des données très désagrégées issues du recensement de la population marocaine de 2004. En effet, la population âgée de 15 ans et plus est répartie en 22 niveaux d'enseignement, et ce pour chacune des 1534 communes. Les données concernant le capital public (les infrastructures) et l'accroissement de la population proviennent toutes du HCP, mais à partir de bases différentes (voir après). Soulignons que l'absence d'un identifiant commun a rendu la tâche difficile. Nous avons été contraints de mener des vérifications minutieuses pour chaque unité géographique pour valider la fusion des différentes bases de données.

La compilation de cette base de données permet de mettre en évidence l'ampleur des inégalités éducatives et des disparités de développement humain à un échelon très fin du territoire. Elle permet aussi de tester l'existence d'autocorrélation spatiale sur des données communales.

Dans un premier temps, nous allons présenter un bref récapitulatif méthodologique sur l'analyse spatiale. Dans un deuxième temps, nous mobilisons les outils de l'Analyse Exploratoire des Données Spatiales (ESDA) afin d'évaluer l'autocorrélation globale et locale de la distribution spatiale du développement et des inégalités d'éducation dans les communes marocaines. Finalement, nous mettons en évidence la relation entre l'indice de Gini de l'éducation et l'indice de développement humain en utilisant un modèle économétrique spatial.

3.3.1. Rappel méthodologique

Nous assistons ces dernières années à des avancées méthodologiques importantes et à un recours croissant à l'économétrie spatiale (voir par exemple Anselin, 2010 ; Schaffar, 2014).

La disponibilité de données au niveau infranational conduit de plus en plus de travaux empiriques à prendre en compte l'autocorrélation spatiale. Cette dernière nous renseigne sur le degré de correspondance entre proximité géographique et similarité des valeurs observées. En d'autres termes, elle nous informe sur le niveau de la non-indépendance entre les valeurs d'unités géographiques rapprochées.

La modélisation des interdépendances entre unités géographiques requiert de déterminer a priori une matrice de poids spatiale prenant en compte le voisinage entre unités géographiques. Il s'agit d'une matrice carrée composée d'autant de lignes et de colonnes qu'il y a d'unités géographiques. Chaque élément de la matrice renseigne sur l'interaction des deux unités géographiques correspondantes en ligne

et en colonne. Pour éviter certains problèmes de singularité de cette matrice, les valeurs de la diagonale sont conventionnellement considérées nulles. On note N le nombre d'unités géographiques et w_{ij} l'élément de la matrice liant les deux unités i et j . La forme de la matrice de poids la plus adaptée à notre problématique est la matrice de contiguïté d'ordre 1 définie comme suit :

$$w_{ij} = \begin{cases} 1, & \text{si les deux unités } i \text{ et } j \text{ ont une frontière commune} \\ 0, & \text{sinon} \end{cases}$$

La matrice de contiguïté est ensuite standardisée : chaque élément de la matrice est divisé par la somme en ligne de tous les éléments, de telle sorte que la somme en ligne devienne égale à 1.

- ***L'autocorrélation spatiale globale***

Pour calculer l'autocorrélation spatiale, la statistique la plus communément utilisée est la statistique de Moran (I). Elle indique le degré d'association linéaire entre la valeur prise par une variable dans une localisation donnée et la moyenne spatialement pondérée des localisations voisines. Plus exactement, c'est le ratio de la covariance entre observations contigües à la variance totale de la variable d'intérêt.

Pour une variable d'intérêt " y " et deux unités géographiques i et j , la statistique de Moran est définie de la façon suivante :

$$I = \frac{N \sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{S_0 \sum_i (y_i - \bar{y})^2} = \frac{N}{S_0} \cdot \frac{z'Wz}{z'z}$$

où y_i est la valeur observée de l'unité géographique i , \bar{y} est la moyenne de la variable considérée, z est le vecteur des N observations en déviation à la moyenne \bar{y} , w_{ij} est l'élément de la matrice de poids spatiale W , et S_0 est la somme des éléments de la matrice W , telle que :

$$S_0 = \sum_i \sum_j w_{ij}$$

L'indice a des valeurs comprises entre -1 et 1. L'autocorrélation spatiale est donc considérée négative ($I < 0$) si les valeurs d'unités géographiques rapprochées sont dissemblables. En revanche elle est positive ($I > 0$) quand ces valeurs sont similaires, qu'elles soient élevées ou faibles. Ainsi une autocorrélation spatiale positive traduit une tendance à la concentration dans l'espace de valeurs faibles ou élevées d'une variable d'intérêt et une autocorrélation spatiale négative signifie que chaque localisation tend à être entourée par des localisations voisines pour lesquelles la variable d'intérêt prend des valeurs très différentes. Enfin, l'absence d'autocorrélation spatiale ($I = 0$) réfère à une distribution spatiale parfaitement aléatoire par rapport à la variable d'intérêt dans le territoire.

- **L'autocorrélation spatiale locale**

Si la statistique I de Moran donne une mesure globale de l'autocorrélation spatiale, les indicateurs locaux d'association, introduits par Anselin en 1995, permettent de détecter le degré de corrélation spatiale au niveau local pour chaque unité géographique. Les indicateurs locaux d'association spatiale connu dans la littérature sous le nom LISA (Local indicators of spatial association) ou l'indice de Moran local est l'indicateur le plus utilisé pour mesurer la concentration spatiale locale. Celui-ci satisfait deux critères essentiels :

- ❖ Le LISA donne une indication sur l'étendue d'un regroupement spatial significatif de valeurs similaires
- ❖ La somme des LISA associés à toutes les observations est proportionnelle à un indicateur global d'association spatiale (Longly et Tobon 2004)

Comme dans le cas de l'indice de Moran global, on peut tester la significativité du degré d'autocorrélation spatiale locale de chacune des unités géographiques. Ainsi, celles présentant des valeurs significatives peuvent être représentées sous formes de cartes.

L'indice de Moran local est défini de la façon suivante :

$$I_i = \frac{(N - 1)(y_i - \bar{y}) \sum_{j=1}^N w_{ij}(y_j - \bar{y})}{\sum_{j=1}^N (y_j - \bar{y})^2}$$

L'autocorrélation spatiale peut être introduite explicitement dans la spécification du modèle. Il existe deux principales spécifications des modèles économétriques spatiaux.

- **Le modèle autorégressif**

$$y_i = \rho W y_i + x_i' \beta + \varepsilon_i \quad i = 1, 2, \dots, n.$$

Dans le modèle autorégressif, l'autocorrélation spatiale est incorporée dans la variable endogène décalée pour la matrice de poids W . Ainsi, l'observation y_i pour la commune i est en partie expliquée par la moyenne des observations des communes voisines tel que $W y_i = \sum_{j \neq i} w_{ij} y_j$.

Avec $|\rho| < 1$ où ρ est le paramètre spatial autorégressif.

- **Le modèle à erreurs spatiales**

$$\begin{cases} y_i = x_i' \beta + \varepsilon_i \\ \varepsilon_i = \lambda W \varepsilon_i + \eta_i \end{cases} \quad i = 1, 2, \dots, n.$$

Dans le modèle à erreurs spatiales, l'autocorrélation spatiale est incorporée dans les termes d'erreurs qui sont supposés suivre un processus autorégressif. Le paramètre λ exprime l'intensité de la corrélation spatiale entre les résidus de la régression $|\lambda| < 1$.

3.3.2. Inégalités spatiales d'éducation et de développement humain au Maroc : approche par l'analyse exploratoire des données spatiales

Le tableau n° 24 présente l'IDH et l'indice de Gini de l'éducation des vingt premières et dernières communes en 2004 classées selon l'ordre croissant des inégalités éducatives. La figure n° 3 présente les cartes de ces deux indices pour la même année 2004. Les écarts entre communes ont atteint 0,63 et 0,66 respectivement pour l'indice de Gini de l'éducation et l'IDH (l'indice de Gini de l'éducation se situe entre 0,34 pour la commune Agdal Riyad et 0,97 pour la commune Anemzi et l'IDH se situe entre 0,82 pour la commune Agdal Riyad et 0,16 pour la commune Anemzi). De même l'écart entre communes pour la moyenne d'années de scolarisation est important. Les vingt dernières communes affichent une moyenne largement inférieure à une année de scolarisation tandis que les vingt premières présentent une moyenne d'années de scolarisation supérieure à 6 années.

Les communes qui s'avèrent être particulièrement inégalitaires quant à l'accès à l'éducation (Anemzi, Oulad M'Hammed, Bouchaouene, Imindounit, Ait Oumdis, Sidi Yahya Ou Youssef, Afalla Issen et Aghbar) sont les communes rurales qui présentent les IDH les plus faibles et qui appartiennent aux provinces les moins développées. Les communes qui sont relativement plus égalitaires (Agdal Riyad, Assoukhour Assawda, El Maarif, Rabat Hassan, Touarga, Agdal Fès, Harhoura et Machouar Casablanca) enregistrent les IDH les plus élevés et appartiennent toutes au milieu urbain⁵³ des provinces développées. Les communes les plus inégalitaires appartiennent aux provinces les plus inégalitaires et les provinces les plus inégalitaires appartiennent aux régions les plus inégalitaires.

Les statistiques de Moran calculées pour l'indice de Gini de l'éducation et l'indice de développement humain sont positives et significatives à 1% (tableau n° 25). Ceci confirme l'existence d'une autocorrélation spatiale positive de ces variables. Ainsi, les communes présentant des niveaux élevés d'inégalités éducatives ont tendance à être situées près d'autres communes ayant également des niveaux d'inégalités élevés, et inversement. De même les communes qui sont moins développées ont tendance à avoisiner celles qui sont aussi moins développées et vice versa.

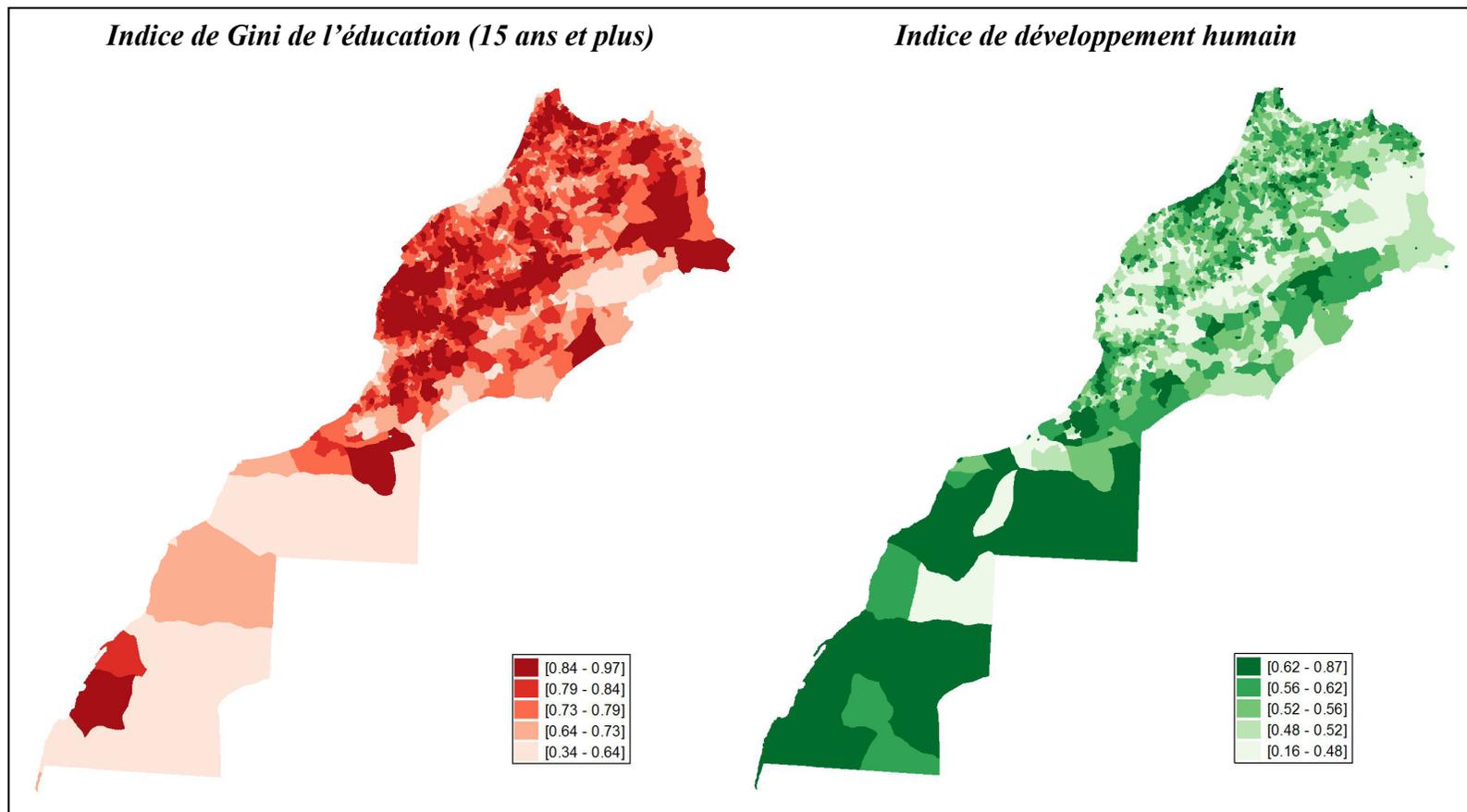
⁵³ Sur les vingt premières communes, les dix-neuf premières sont urbaines et la vingtième est rurale. Il s'agit de la commune de Tifariti appartenant aux régions du Sud.

Tableau 24: Indice de Gini de l'éducation et IDH des vingt premières et dernières communes, 2004

Les vingt premières communes					Les vingt dernières communes				
Commune	IDH	Gini _{edu}	Province d'appartenance	Milieu	Commune	IDH	Gini _{edu}	Province d'appartenance	Milieu
Agdal Riyad	0,82	0,34	Rabat	Urbain	M'Ramer	0,43	0,92	Essaouira	Rural
Assoukhour Assawda	0,78	0,38	Casablanca	Urbain	Takoucht	0,43	0,92	Essaouira	Rural
El Maarif	0,80	0,39	Casablanca	Urbain	Kouzemt	0,45	0,93	Chichaoua	Rural
Rabat Hassan	0,78	0,39	Rabat	Urbain	Adassil	0,42	0,93	Chichaoua	Rural
Touarga	0,76	0,40	Rabat	Urbain	Tizi N'Ghachou	0,40	0,93	Midelt	Rural
Agdal	0,75	0,40	Fès	Urbain	Oulad Ali Mansour	0,40	0,93	Tétouan	Rural
Harhoura	0,72	0,41	Témara	Urbain	Tahelouante	0,38	0,93	Essaouira	Rural
Machouar Casablanca	0,77	0,42	Casablanca	Urbain	Bouabout Amdlane	0,38	0,93	Chichaoua	Rural
Neima	0,70	0,42	Oujda	Urbain	Sidi Boukhalif	0,37	0,93	Azilal	Rural
Anfa	0,80	0,43	Casablanca	Urbain	Ait Blal	0,34	0,93	Azilal	Rural
El Jadida	0,76	0,43	El Jadida	Urbain	Assais	0,41	0,94	Essaouira	Rural
Sidi Bernoussi	0,76	0,43	Casablanca	Urbain	Boutferda	0,40	0,94	Beni Mellal	Rural
Ain Chock	0,76	0,43	Casablanca	Urbain	Aghbar	0,35	0,94	Al Haouz	Rural
Gueliz	0,75	0,43	Marrakech	Urbain	Afalla Issen	0,40	0,95	Chichaoua	Rural
Errachidia	0,72	0,43	Errachidia	Urbain	Sidi Yahya Ou Youssef	0,36	0,95	Midelt	Rural
Souissi	0,77	0,44	Rabat	Urbain	Ait Oumdis	0,34	0,95	Azilal	Rural
Sidi Belyout	0,76	0,44	Casablanca	Urbain	Imindounit	0,30	0,95	Chichaoua	Rural
Ain Sebaa	0,75	0,44	Casablanca	Urbain	Bouchaouene	0,29	0,96	Figuig	Rural
Hay Hassani	0,76	0,45	Casablanca	Urbain	Oulad M'Hammed	0,32	0,97	Taourirt	Rural
Tifariti	0,75	0,45	Es-smara	Rural	Anemzi	0,16	0,97	Midelt	Rural

Source : Calculs de l'auteur à partir des données du recensement de la population du HCP, 2004.

Figure 3 : Indice de Gini de l'éducation et de l'IDH dans les communes marocaines, 2004



Source : Auteur.

Tableau 25 : Statistique de Moran pour l'IDH et l'indice de Gini de l'éducation

Variable	I-Moran	E(I)	Ecart-Type	moyenne	Probabilité
Gini ^{edu}	0.531554	-0.0007	0.0161	-0.0013	0.001
IDH	0,475563	-0,0007	0,016	-0,0007	0,001

Source : Auteur.

Toutefois, l'analyse spatiale globale peut dissimuler certaines spécificités locales. D'où, l'intérêt d'identifier les regroupements spatiaux tout en examinant leur instabilité locale.

Ainsi, l'analyse de l'autocorrélation spatiale locale est étudiée à l'aide de deux outils: le diagramme de Moran et les statistiques LISA, (Local Indicators of Spatial Association). Le premier outil permet de visualiser les schémas de localisation associés à des unités spatiales (Anselin, 1996). En abscisse, est représentée la valeur standardisée de la variable et en ordonnée son décalage spatial standardisé. Chaque quadrant correspond à un type différent d'association spatiale locale existant entre une commune et son voisinage selon que la valeur est élevée (au-dessus de la moyenne) ou faible (au-dessous de la moyenne) : HH (Elevé - Elevé) ; LH (Faible - Elevé), HL (Elevé - Faible) et LL (Faible-Faible).

- HH : une commune associée à une valeur élevée entourée de communes associées à des valeurs élevées.
- LH : une commune associée à une valeur faible entourée de communes associées à des valeurs élevées.
- HL : une commune associée à une valeur élevée entourée de communes associées à des valeurs faibles.
- LL : une commune associée à une valeur faible entourée de communes associées à des valeurs faibles.

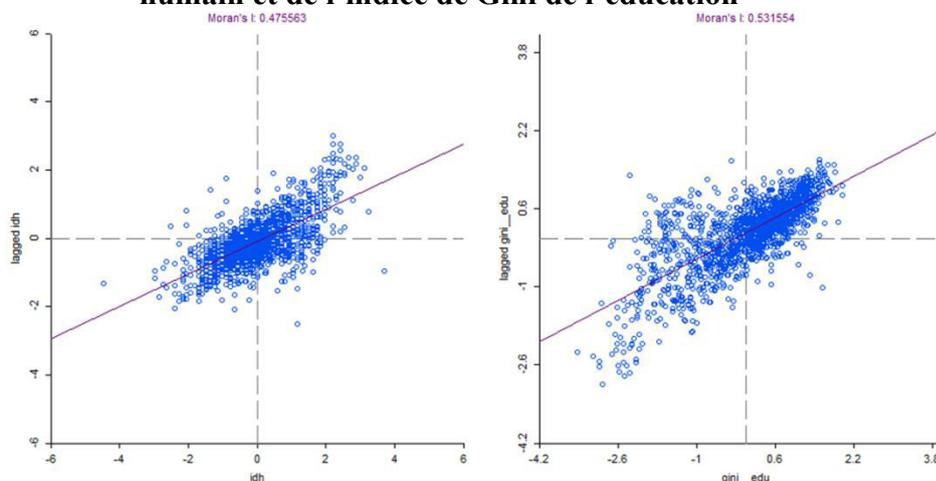
Le premier et le dernier quadrant représentent une association spatiale positive, Les deux autres représentent une association spatiale négative. Les diagrammes de Moran utilisés dans cette étude sont construits à partir de la matrice de contiguïté d'ordre 1.

Afin de visualiser et distinguer le regroupement d'observations similaires ou dissemblables selon les quatre quadrants d'association spatiale, nous présentons dans le graphique n° 25 le diagramme de Moran de l'indice de développement humain et de l'indice de Gini de l'éducation.

En effet, 71% des communes marocaines se caractérisent par une association spatiale positive quant à la répartition spatiale du développement humain. 28% des communes se trouvent dans le quadrant HH et 43% se trouvent dans le quadrant LL. Les 29% des communes restantes se caractérisent par une association spatiale négative où 17% se regroupent dans le quadrant HL « îlot de richesse » et 12% se concentrent dans le quadrant LH « mouton noir » (tableau n° 26). En ce qui concerne l'indice de Gini de l'éducation, le diagramme de Moran montre que 79% des communes relèvent d'une association spatiale positive dont 51% se concentrent dans le quadrant HH et 29% se regroupent dans le quadrant LL. Par contre, 21% des communes se caractérisent par une association spatiale de

valeurs dissimilaires (13% dans le quadrant LH et 8% dans le quadrant HL) (tableau n° 26).

Graphique 25: Diagramme de Moran de l'indice de développement humain et de l'indice de Gini de l'éducation



Source : Auteur.

Si le diagramme de Moran permet d'identifier l'existence de l'hétérogénéité spatiale entre les communes marocaines, il ne fournit toutefois, aucune information sur la significativité des schémas d'association spatiale. Ces derniers doivent être approuvés par le recours aux statistiques LISA qui constituent le second outil d'analyse de l'autocorrélation spatiale locale.

Tableau 26 : Diagramme de Moran : récapitulation des associations spatiales (en %)

Indicateur	Association positive			Association négative		
	HH	LL	Total	HL	LH	Total
IDH	28,27	43,02	71,29	16,95	11,75	28,7
Gini ^{edu}	50,8	28,63	79,43	7,87	12,68	20,55

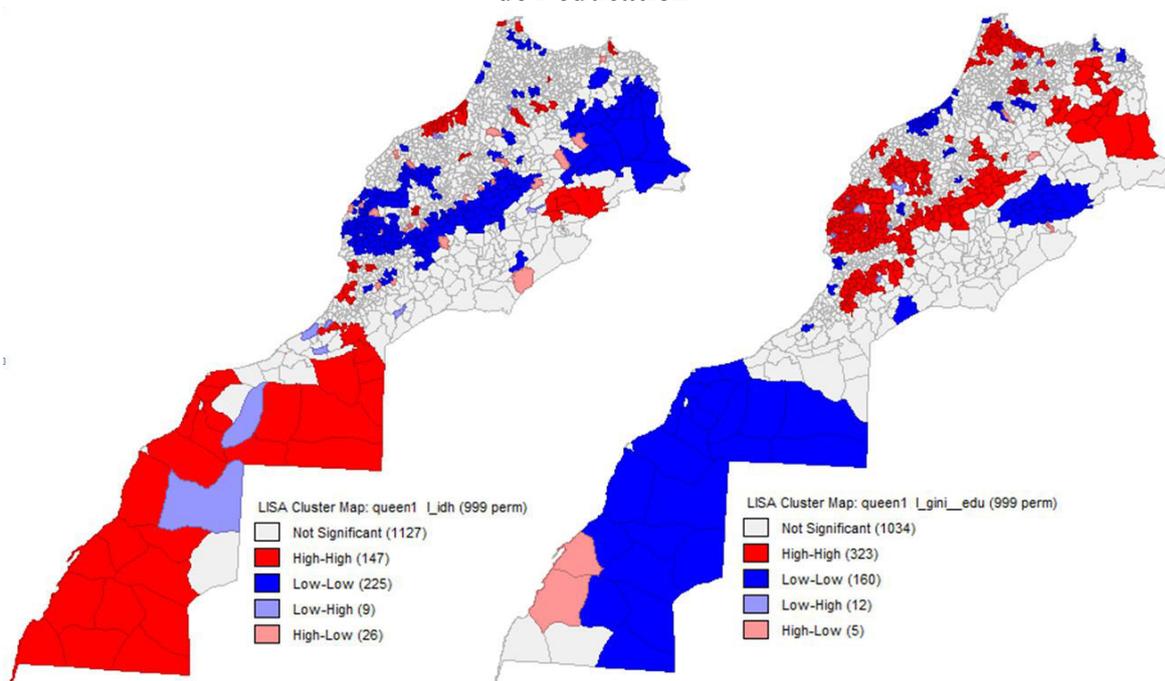
Source : Auteur.

Après avoir localisé les groupements des associations spatiales des communes, nous spécifions celles qui sont significatives en utilisant les statistiques LISA. La figure n° 4 présente les cartes de significativité de Moran (au niveau de 5%) pour l'indice de développement humain et l'indice de Gini de l'éducation. Une tendance globale à l'autocorrélation spatiale positive est confirmée. En effet, pour l'indice de développement humain et l'indice de Gini de l'éducation, 92% (respectivement 97%) des statistiques LISA se situent dans les quadrants HH ou LL.

Sur la base des statistiques LISA, nous avons pu identifier deux groupes de communes. Le premier est constitué de communes ayant des niveaux élevés d'inégalité et le deuxième est composé de communes ayant de faibles niveaux d'inégalité. Dans le premier groupe, on distingue trois grappes de communes identifiées sur la base d'une association positive : une grappe de plus de 200 communes appartenant aux provinces d'Essaouira, Chichaoua, El Haouz, Rhamna, Safi, Chtouka-Aït Baha, Taroudant, Tiznit, Azilal et Midelt; une deuxième grappe de 60 communes localisées au Nord dans les provinces de Chefchaouene, Larache,

Tetouan, Anjra, Taounate, Taza et Sidi Kacem; une troisième grappe concentrée à l’Oriental dans les provinces de Berkane, Jerada, Figuig et Taourirt.

Figure 4 : LISA pour l’indice de développement humain et l’indice de Gini de l’éducation



Source : Auteur.

Le second groupe est aussi constitué de trois grappes: la première est composée de 49 communes dans les provinces de Mohammedia, Casablanca, Rabat, Salé et Témara; la deuxième comprend 25 communes appartenant aux provinces du Sud (Aousserd, Dakhla, Boujdour, Laayoune, Es-smara et Tarfaya); la troisième est constituée de 19 communes du nord de la province d’Errachidia.

En plus des six grappes de communes identifiées, 13 petits regroupements HH significatifs sont mis en évidence dont chacun comporte une dizaine de communes qui sont pour la plupart voisine d’une municipalité ou d’un arrondissement (à titre d’exemple les municipalités d’Agadir, Khouribga, Nador, Meknès, Machouar Fès El Jadid et Martil et les arrondissements de Marrakech Medina et de Tanger).

La cartographie de l’autocorrélation spatiale locale réaffirme bien que le développement humain et les inégalités éducatives sont fortement localisés, et nous informe aussi sur la grande similitude entre la distribution spatiale de ces derniers. En effet, les grappes de communes appartenant au quadrant HH pour l’indice de Gini de l’éducation tendent à recouper celles appartenant au quadrant LL pour l’IDH et vice versa. Il semblerait ainsi que la localisation géographique constitue un facteur discriminant de la scolarisation puisqu’elle conditionne l’accès équitable à une éducation pour tous, qui conditionne à son tour le niveau de développement humain. Afin de vérifier ce constat nous menons dans ce qui suit une modélisation par l’économétrie spatiale.

3.3.3. Les effets des inégalités d'éducation sur le développement humain

Cette analogie observée entre les deux cartes communales nous incite à prolonger l'analyse en examinant, à travers un modèle économétrique spatial en coupe transversale, l'effet des inégalités d'éducation sur le développement humain. Cette approche a été peu explorée jusqu'à notre connaissance, à l'exception de Umar et al. (2014b) qui ont analysé le cas du Nigéria, aucune étude n'a utilisé une modélisation spatiale pour analyser le lien de causalité entre les inégalités éducatives et le développement territorial.

3.3.3.1. Définition des variables

Nous utilisons donc l'IDH comme variable expliquée. Comme l'IDH tient compte uniquement des réalisations moyennes⁵⁴, l'inclusion d'un indice sur la distribution de l'une de ses composantes comme variable explicative ne pose pas de problème.

Toutefois, en guise de contrôle et dans le but d'affirmer la robustesse de nos résultats nous calculons aussi un "IDH corrigé" au niveau communal qui comprend uniquement les deux composantes santé et niveau de vie. En effet, en plus de l'IDH, l'IDH corrigé nous aidera à discerner "l'effet pur" des inégalités éducatives sur le développement de ces deux dimensions tout en évitant toute possible corrélation préalable entre la composante éducation de l'IDH et sa distribution.

Ainsi, en l'absence de l'espérance de vie à la naissance, du PIB et du revenu au niveau communal, les composantes santé et niveau de vie de l'IDH corrigé sont appréhendées respectivement à travers le taux de survie des enfants de moins d'un an⁵⁵ et les dépenses annuelles moyennes par ménage. Cette approche semble être plus adéquate au niveau communal.

Notre modèle est estimé en coupe transversale sur 1534 communes pour l'année 2004. Nous considérons outre *GiniEdu*, comme autres variables explicatives, l'accès aux infrastructures de base par les ménages (*Infra*) et l'accroissement de la population (*PopGrowth*).

Le modèle est défini comme suit :

$$\begin{cases} IDH_i = \beta_0 + \beta_1 GiniEdu_i + \gamma_1 Infra_i + \gamma_2 PopGrowth_i + \varepsilon_i \\ \text{où } \varepsilon_i \sim N(0, \sigma^2) \quad i = 1 \dots 1534. \end{cases} \quad (S1)$$

À défaut de disposer d'une mesure du capital public au niveau communal, nous avons l'avons approché par un indice synthétique relatif au degré d'accessibilité de la population aux infrastructures de base issue du HCP pour

⁵⁴ C'est la raison pour laquelle le PNUD a introduit pour la première fois en 2010 l'IDH « ajusté aux inégalités » (IDHI) qui permet de rendre compte non seulement des réalisations moyennes d'un pays en matière de santé, d'éducation et de revenu, mais également des niveaux d'inégalités dans ces dimensions au sein de la population.

⁵⁵ Signalons que ce taux est utilisé par l'Organisation Mondiale de la Santé comme l'une des variables indiquant l'état de santé de la population d'un pays donné. Ce dernier est basé sur la probabilité pour qu'un enfant né vivant décède avant son premier anniversaire.

l'année 2004. Cet indice comporte trois composantes : (i) le degré d'accessibilité au réseau routier, apprécié par le ratio de population vivant à moins de deux kilomètres d'une route praticable en toutes saisons; (ii) le degré d'accessibilité à l'eau potable mesuré par la part des ménages ayant accès à une source d'approvisionnement public; (iii) le degré d'électrification, évalué par la part de logements électrifiés. L'indice « *infra* » est calculé par la moyenne arithmétique équipondérée des taux des trois composantes (accessibilité à l'eau, à l'électricité et à la route).

L'accroissement de la population est calculé sur une période de dix années, entre 1994 et 2004 à partir de données sur la population par commune fourni par le HCP. Cette variable peut être considérée comme représentant les besoins à satisfaire en termes de santé, éducation et niveau de vie.

L'estimation de l'équation (1) par la méthode des moindres carrés ordinaires (MCO) rejette les hypothèses de normalité et d'homoscédasticité des termes d'erreur. En effet, le test de normalité de Jarque-Bera conçu sur la base des coefficients d'asymétrie et d'aplatissement de la distribution rejette l'hypothèse de normalité. La statistique de Jarque-Bera pour les modèles⁵⁶ de l'IDH et de l'IDH corrigé est respectivement égale à 1099 ($p=0.00000$) et 1325 ($p=0.00000$). Le test de Breush-Pagan basé sur la comparaison entre l'estimation par les MCO et celle avec le maximum de vraisemblance rejette l'hypothèse d'homoscédasticité des résidus. La statistique de Breush-Pagan pour les modèles de l'IDH et de l'IDH corrigé est respectivement égale à 264 ($p=0.00000$) et 298 ($p=0.00000$).

Bien évidemment, le test de Moran global présenté auparavant dans l'analyse exploratoire des données spatiales ne concernait pas un modèle économétrique spécifique. Néanmoins, ce test peut être ajusté pour examiner l'autocorrélation spatiale des termes d'erreur découlant de l'estimation de l'équation (1) par la méthode des MCO. Il suffit de remplacer le vecteur z ⁵⁷ dans l'écriture matricielle de la statistique du Moran global par le vecteur des termes d'erreur estimés $\hat{\varepsilon}$.

Pour parvenir à une bonne spécification du modèle, nous commençons par calculer la statistique de Moran. Le test appliqué aux deux modèles de l'équation (1) estimés par MCO confirme l'existence d'autocorrélation spatiale des résidus. En effet, la statistique de Moran pour les résidus du modèle de l'IDH et de l'IDH corrigé est respectivement égale à 0.249 ($p=0.00000$) et 0.214 ($p=0.00000$).

Si le test de Moran confirme l'existence d'une autocorrélation spatiale des termes d'erreur, il ne permet pas cependant de déterminer la spécification du modèle spatial la mieux adéquate. Les deux modèles les plus fréquemment utilisés sont le modèle autorégressif et le modèle à erreurs spatial.

Afin de discriminer entre ces deux types de modèles et de spécifier la forme la plus adaptée de l'autocorrélation spatiale nous faisons appel aux deux tests du multiplicateur de Lagrange, ainsi que de leurs versions robustes (Anselin et Florax, 1995 ; Anselin, 1996): LM-Lag et RLM-Lag pour une variable (endogène ou

⁵⁶ Les deux modèles ont respectivement l'IDH et l'IDH corrigé comme variable dépendante

⁵⁷ Le vecteur z dans l'écriture matricielle représentait les N observations en déviation à la moyenne \bar{Y} .

exogène) décalée, et LM-Err et RLM-Err pour une autocorrélation spatiale des erreurs. Ainsi la statistique RLM-Lag pour le modèle de l'IDH et de l'IDH corrigé est respectivement égale à 0.82 ($p=0.36397$) et 0.03 ($p=0.86907$) tandis que la statistique RLM-Err pour le modèle de l'IDH et de l'IDH corrigé est respectivement égale à 167.09 ($p=0.00000$) et 71.48 ($p=0.00000$).

En utilisant les tests robustes de spécification du multiplicateur de Lagrange, nous aboutissons ainsi à rejeter la présence d'une variable endogène ou exogène spatialement décalée et à retenir, avec une probabilité marginale largement inférieure à 1 %, le modèle à autocorrélation spatiale des erreurs comme étant le plus adapté à notre cas. Différents types de processus sont évoqués dans la littérature, toutefois la spécification autorégressive spatiale est la plus utilisée. Le modèle s'écrit alors de la façon suivante :

$$\begin{cases} IDH_i = \beta_0 + \beta_1 GiniEdu_i + \gamma_1 Infra_i + \gamma_2 PopGrowth_i + \varepsilon_i \\ \text{où } \varepsilon_i = \lambda W\varepsilon_i + \eta_i \quad i = 1 \dots 1534. \end{cases} \quad (S2)$$

3.3.3.2. *Un modèle économétrique spatial*

Le tableau n° 27 fournit les résultats des estimations des équations (S1) et (S2) pour chacune des deux variables dépendantes (l'IDH et l'IDH corrigé). Le R^2 n'est pas approprié pour la comparaison entre le modèle MCO, le modèle à variable décalée ou le modèle à autocorrélation spatiale des erreurs. Il faut confronter les valeurs du logarithme de vraisemblance et par la suite comparer entre les valeurs de chacun des deux critères d'information (Akaike et Schwartz). Ainsi le logarithme de vraisemblance est nettement supérieur pour les modèles à autocorrélation spatiale des erreurs par rapport aux modèles classiques. De même les deux critères d'information d'Akaike et de Schwartz sont nettement inférieurs pour les modèles d'erreur spatial par rapport aux modèles classiques.

Le paramètre λ présente une intensité de corrélation spatiale positive entre les termes d'erreur. Il est de l'ordre de 0.42 et 0.37 respectivement pour le modèle de l'IDH et de l'IDH corrigé, et ce avec une probabilité marginale largement inférieure à 1% pour les deux modèles. Ce constat rejoint ainsi celui dressé par le test du ratio de vraisemblance pour les deux modèles et atteste de l'ampleur des interdépendances spatiales entre les résidus.

Les deux premières colonnes du tableau n° 27 présentent les résultats du modèle MCO et du modèle à autocorrélation spatiale des erreurs avec l'IDH comme variable dépendante tandis que les deux dernières concernent les résultats de ces deux modèles avec l'IDH corrigé comme variable dépendante. Les modèles avec l'IDH ou l'IDH corrigé comme variable dépendante présentent le signe négatif attendu pour le coefficient (β_1) de l'indice de Gini de l'éducation, et ce aussi bien pour le MCO que pour le modèle d'erreur spatial. Par conséquent, le faible niveau de développement humain est associé à des niveaux d'inégalité d'éducation plus prononcées. Le constat dégagé par l'ESDA est confirmé par le modèle économétrique spatial, qui atteste d'un lien de causalité en contrôlant toute éventuelle corrélation entre la composante éducation de l'IDH et l'indice de Gini de l'éducation.

Tableau 27 : Résultats des estimations du modèle par les MCO et du modèle à autocorrélation spatiale des termes d'erreur

Variables explicatives	Variable dépendante : IDH		Variable dépendante : IDH corrigé	
	Modèle classique	Modèle d'erreur spatial	Modèle classique	Modèle d'erreur spatial
Constante	0.784 (25.181)***	0.832 (23.690)***	0.437 (14.43)***	0.48 (14.049)***
Gini_Edu	-0.586 (-47.26)***	- 0.600 (-45.70)***	-0.314 (-26.07)***	- 0.324 (-25.23)***
Infra	0.016 (2.869)***	0.021 (3.535)***	0.013 (2.394)**	0.018 (3.137)***
Pop_Growth	-0.091 (-7.77)***	-0.073 (-5.294)***	-0.115 (-9.98)***	-0.097 (-7.262)***
Lambda	-	0.416 (12.763)***	-	0.373 (11.092)***
Nb d'observations	1534	1534	1534	1534
F-Statistic	(1700.9)***	-	(671.2)***	-
R ²	0.769	0.800	0.568	0.614
Log de Vraisemblance	2692.25	2778.02	2733.37	2798.86
Test du ratio de vraisemblance	-	(171.55)***	-	(130.99)***
Critère d'information Akaike	-5376.50	-5548.05	-5458.73	-5589.72
Critère de Schwartz	-5355.15	-5526.70	-5437.39	-5568.38
Moran	0.249 (15.616)***	-	0.214 (13.489)***	-
LM-Lag	(72.635)***	-	(106.44)***	-
RLM-Lag	(0.824)	-	(0.02)	-
LM-Err	(238.9)***	-	(177.9)***	-
RLM-Err	(167.1)***	-	(71.48)***	-

Source : Auteur.

Les chiffres entre parenthèses correspondent aux statistiques de Student.

***, ** et * indiquent que les variables sont respectivement significatives à 1%, 5 % et 10%

LM-Lag et RLM-Lag renvoient respectivement au test du multiplicateur de Lagrange appliqué sur la variable endogène décalée et à sa version robuste. LM-Err et RLM-Err désignent respectivement le test du multiplicateur de Lagrange appliqué sur les termes d'erreur et sa version robuste.

Les résultats de la modélisation économétrique spatiale sont conformes également à nos attentes concernant le signe des variables de contrôle. Ainsi le capital public d'une commune a un impact positif sur son développement humain. Plus la commune assure une accessibilité aux infrastructures de base, plus elle a tendance à avoir un IDH élevé. Par contre, l'accroissement de la population des communes marocaines a un impact négatif sur leur développement territorial puisqu'il renvoie à l'accroissement des besoins à satisfaire lié d'une certaine manière à la démographie.

3.3.3.3. *Un modèle polynomial spatial*

Dans l'analyse au niveau communal, il reste un point important à traiter, et qui nous a préoccupé tout au long du chapitre II, c'est de tester l'existence d'un effet différencié selon le niveau d'inégalité éducative sur l'IDH. Afin de tester la possibilité d'un éventuel impact non linéaire des inégalités éducatives sur le développement territorial, nous introduisons les termes quadratique et cubique de la variable de Gini de l'éducation dans les régressions précédentes. Le modèle s'écrit de la façon suivante :

$$\left\{ \begin{array}{l} y_i = \sum_{k=0}^{K+1} \beta_k (GiniEdu_i)^k + \gamma_1 Kphysique_i + \gamma_2 PopGrowth_i + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2) \quad i = 1 \dots 1534 . \\ \text{où } y_i = IDH_i \quad \text{ou} \quad y_i = IDH_i^{corr} \end{array} \right. \quad (3)$$

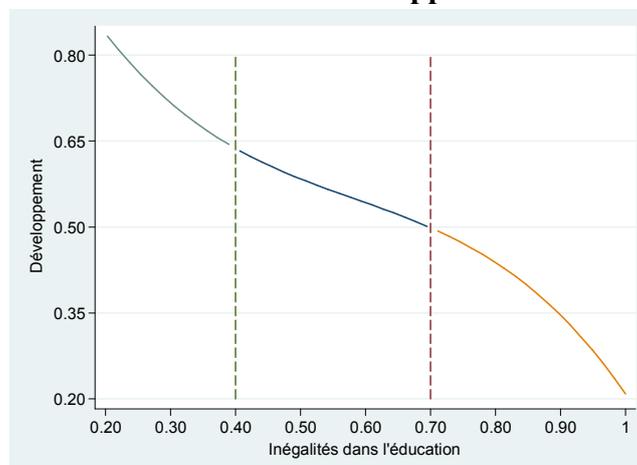
$$\left\{ \begin{array}{l} y_i = \sum_{k=0}^{K+1} \beta_k (GiniEdu_i)^k + \gamma_1 Kphysique_i + \gamma_2 PopGrowth_i + \varepsilon_i \\ \varepsilon_i = \lambda W \varepsilon_i + \eta_i \quad i = 1 \dots 1534 . \\ \text{où } y_i = IDH_i \quad \text{ou} \quad y_i = IDH_i^{corr} \end{array} \right. \quad (4)$$

Le tableau n° 28 fournit les résultats des estimations des spécifications quadratiques { (3.1) et (4.1) } et cubiques { (3.2) et (4.2) } respectivement du modèle MCO et du modèle à autocorrélation spatiale des erreurs et ce pour chacune des deux variables dépendantes (l'IDH et l'IDH corrigé).

Les estimations des modèles ayant l'IDH comme variable dépendante montrent que les termes linéaire, quadratique et cubique de l'indice de Gini de l'éducation sont tous significatifs. Ainsi pour la spécification cubique, les termes linéaire et cubique ont un signe négatif alors que le terme quadratique a un signe positif. De même, les estimations des modèles ayant l'IDH corrigé comme variable dépendante présentent des termes linéaire et cubique tous négatifs et significatifs, tandis que le terme quadratique est significatif uniquement pour la spécification cubique. Ainsi cette dernière semble être plus appropriée.

Le graphique n° 26 schématise les différentes étapes de la relation selon la spécification cubique retenue. Lorsque le niveau d'inégalité éducative est très élevé (indice de Gini proche de 1) la fonction a tendance à être concave. Cette concavité diminue au fur et à mesure que l'inégalité baisse jusqu'à atteindre un premier seuil de 0.70. En deçà de ce niveau, la fonction tend à suivre la même pente jusqu'à un niveau d'inégalité de 0.40 ; à partir de ce seuil la fonction commence à devenir convexe. L'impact d'une réduction des inégalités d'éducation sur le développement humain apparaît donc de manière générale, mais cet impact est moins prononcé pour les communes qui présentent des niveaux d'inégalités élevés.

Graphique 26 : Illustration de la trajectoire du lien entre inégalités dans l'éducation et développement



Source : Auteur.

Tableau 28: Résultats des estimations du modèle par les MCO et du modèle à autocorrélation spatiale des termes d'erreur

Variable dépendante	Indice de développement humain				Indice de développement humain corrigé			
	Modèle classique		Modèle d'erreur spatial		Modèle classique		Modèle d'erreur spatial	
Variables explicatives	Quadratique	Cubique	Quadratique	Cubique	Quadratique	Cubique	Quadratique	Cubique
Constante	0.714 (16.546)***	1.237 (8.688)***	0.766 (16.233)***	1.141 (7.748)***	0.403 (9.59)***	0.880 (6.346)***	0.458 (9.919)***	0.806 (5.574)***
Gini_Edu	-0.360 (-3.702)***	-2.728 (-4.387)***	-0.385 (-3.674)***	-2.072 (-3.257)***	-0.204 (-2.153)**	-2.362 (-3.9)***	-0.251 (-2.452)**	-1.823 (-2.91)***
Gini_Edu ²	-0.165 (-2.337)*	3.359 (3.664)***	-0.156 (-2.063)**	2.346 (2.512)**	-0.080 (-1.174)	3.131 (3.506)***	-0.053 (-0.712)	2.28 (2.481)*
Gini_Edu ³	- (-3.855)***	-1.706 (-3.855)***	- (-2.689)***	-1.210 (-2.689)***	- (-3.607)***	-1.556 (-3.607)***	- (-3.607)***	-1.129 (-2.548)*
K_Physique	0.011 (1.931)*	0.012 (2.001)**	0.016 (2.657)***	0.016 (2.664)***	0.011 (1.866)*	0.011 (1.927)*	0.0169 (2.724)***	0.016 (2.724)***
Pop_Growth	-0.090 (-7.680)***	-0.087 (-7.368)***	-0.71 (-5.076)***	-0.068 (-5.00)***	-0.11 (-9.93)***	-0.111 (-9.636)***	-0.0966 (-7.215)***	-0.094 (-7.05)***
Lambda	- (12.703)***	- (12.463)***	0.414 (12.703)***	0.408 (12.463)***	- (11.06)***	- (11.06)***	0.373 (11.06)***	0.367 (10.82)***
Nb d'observations	1534	1534	1534	1534	1534	1534	1534	1534
F-Statistic	(1280.8)***	(1036.9)***	-	-	(503.9)***	(408.8)***	-	-
R ²	0.770	0.772	0.801	0.801	0.568	0.572	0.614	0.615
Log de Vraisemblance	2694.98	2702.41	2780.15	2783.73	2734.06	2740.57	2799.13	2802.35
Test du ratio de vraisemblance	-	-	(170.33)***	(162.65)***	-	-	(130.14)***	(123.57)***
Critère d'information Akaike	-5379.97	--5392.82	-5550.3	-5555.47	-5458.13	-5469.14	-5588.26	-5592.7
Critère de Schwartz	-5353.29	-5360.81	-5523.62	-5523.46	-5431.45	-5437.12	-5561.58	-5560.69
Moran	0.248 (15.5946)***	0.24 (15.144)***	-	-	0.214 (13.460)***	0.206 (13.027)***	-	-
LM-Lag	(75.707)***	(65.175)***	-	-	(108.02)***	(97.396)***	-	-
RLM-Lag	(0.451)	(1.337)	-	-	(0.102)	(0.0002)	-	-
LM-Err	(237.426)***	(223.276)***	-	-	(176.426)***	(164.75)***	-	-
RLM-Err	(162.172)***	(159.437)***	-	-	(68.54)***	(67.35)***	-	-

Source : Auteur.

Les chiffres entre parenthèses correspondent aux statistiques de Student.

***, ** et * indiquent que les variables sont respectivement significatives à 1%, 5 % et 10%

LM-Lag et RLM-Lag renvoient respectivement au test du multiplicateur de Lagrange appliqué sur la variable endogène décalée et à sa version robuste. LM-Err et

RLM-Err désignent respectivement le test du multiplicateur de Lagrange appliqué sur les termes d'erreur et sa version robuste.

3.4. Inégalités éducatives et développement économique des provinces marocaines

Cette section aborde un autre échelon du territoire. Il s'agit de la province qui représente un niveau intermédiaire entre la commune et la région au Maroc. Notre échantillon est composé de 61 provinces durant la période 1998-2013 (976 observations). La province est la plus petite unité géographique du territoire marocain qui permette la compilation de données sur la période considérée des variables mesurant le développement local et les inégalités d'éducation.

Dans un premier temps, nous faisons une revue de la littérature empirique de la relation entre les inégalités éducatives et le développement économique au niveau infranational. Puis, nous définissons les variables et le modèle économétrique utilisés. Enfin, nous présentons et commentons les résultats obtenus.

3.4.1. Les enseignements de la littérature au niveau territorial

Une série de travaux menée sur des pays asiatiques met en évidence au niveau infranational une relation négative entre les inégalités éducatives et le développement économique. Rao et Jani (2008), Zhang et Jin (2010) et Ilon (2011) ont respectivement analysé le cas de la Malaisie pour la période 1986-2005, de la Chine pour la période 1978-2004 et de la Corée pour la période 1965-2004. Parmi les travaux réalisés en données de panel, on peut notamment citer Digdowiseiso (2009) qui a étudié le cas de 23 provinces indonésiennes pour la période 1996-2005, Hassan et Mirza (2007) qui ont analysé le cas de provinces pakistanaises sur la période 1973-1998, Castelló et Mukhopadhyay (2013) qui ont traité le cas de 16 États indiens pour la période 1961-2001 et Yang et Li (2007) qui ont étudié le cas de 31 provinces de la Chine sur la période 1996-2004. Leurs conclusions coïncident pour affirmer qu'une baisse de l'indice de Gini de l'éducation impacte positivement le développement.

Une autre série de travaux concernant les pays de l'Europe de l'Ouest indiquent plutôt que les inégalités initiales d'éducation stimulent la croissance économique, en s'appuyant principalement sur des estimations en coupe transversale. Il s'agit notamment de Simões et Duarte (2010) et Rodríguez-Pose et Tselios (2010) qui ont respectivement étudié le cas de 30 régions portugaises et de 102 régions ouest-européennes.

Trois articles seulement ont analysé la non-linéarité de la relation entre inégalités éducatives et développement économique au sein d'un même pays. Gille (2015) qui analyse un panel de 29 États indiens pour la période 1987-2010 justifie, en introduisant un terme d'interaction⁵⁸, que la relation est non linéaire et qu'elle est positive pour les États les plus riches et négative pour les États relativement pauvres. L'auteur souligne "The relation between human capital distribution and income per capita is not linear and depends on the level of development ...Therefore, it seems that in States with a low income per capita the education equality index is positively correlated with income per capita, whereas in States

⁵⁸ Le terme d'interaction utilisé, qui est l'indice d'égalité d'éducation multiplié par une variable indicatrice du niveau de développement, apparaît significatif à 1%.

with a higher income per capita the relation is significantly negative. This result provides a key for reading the results from previous literature".

De même, Yu et al. (2015) explorent la non linéarité de la relation à travers un panel de 31 provinces chinoises pour la période 1990-2010 en introduisant des termes quadratique et cubique de l'indice de Gini de l'éducation. Ils rejoignent ce qu'a trouvé Gille (2015) et rapportent que la relation dépend du niveau de développement de la province. "An important finding that complements the findings of previous research is the non-linear relationship between inequality in educational attainment and economic growth in China. There appears to be a different inequality in educational attainment growth nexus according to the level of a region's economic development in Chinese provinces (Yu et al., 2015)".

Si Gille (2015) et Yu et al. (2015) proposent un effet différencié selon le niveau de développement au sein du même pays, Güngör (2010) analysant un panel de 67 provinces de la Turquie pour la période 1975-2000 suggère quant à lui un effet de seuil du niveau d'inégalité d'éducation. Ce niveau au-delà duquel l'amélioration de la distribution de l'éducation de certaines provinces conduit à des taux de croissance plus élevés n'est pas explicitement déterminé par Güngör (2010). Toutefois son existence est confirmée par l'auteur en introduisant des termes quadratique, cubique et quartique de l'indice de Gini de l'éducation. The results provide confirmation of a nonlinear relationship between economic growth and education inequality. Economic growth is negatively related to education inequality at low levels of the education Gini coefficient, indicating that a more even distribution of education attainment leads to higher growth rates for provinces with more even education distributions. For provinces with relatively high levels of education inequality, on the other hand, a positive association exists between growth and education inequality (Güngör, 2010).

Toutefois, dans les articles de Gille (2015), Yu et al. (2015), Güngör (2010), aucun n'a analysé la non linéarité de la relation au niveau infranational moyennant un modèle non paramétrique. L'étude de cas des provinces marocaines, pour lesquelles nous avons compilé des données annuelles sur 15 années, nous permet de mener une modélisation non paramétrique et semi-paramétrique en données de panel.

3.4.2. Définition des variables et spécification du modèle

Notre échantillon est composé de 61 provinces marocaines pour la période 1998-2013. La variable dépendante mesurant le développement local est le logarithme népérien du PIB par habitant de la province. Les données relatives au PIB de chaque province entre 1998 et 2013 proviennent de la Direction des Etudes et Prévisions Financières (DEPF) auprès du Ministère de l'économie et des finances. Le PIB par habitant de chaque province est calculé en rapportant le PIB sur la population de la province puis converti en PPA \$ constants 2005 en suivant la même démarche utilisée dans la section 1.3 de ce chapitre. Les données de la population utilisées sont issues du HCP.

Il y a lieu de souligner que l'une des principales difficultés rencontrées lorsqu'il s'agit d'évaluer les contributions provinciales à la richesse nationale est

l'harmonisation entre agrégats provinciaux et nationaux et que l'évaluation des PIB provinciaux est soumise à différentes hypothèses⁵⁹.

Le modèle général se présente pour chaque province (i) à l'année (t) sous la forme suivante :

$$\begin{cases} y_{it} = \beta_0 + \beta_1 IneqEdu_{it-1} + \gamma_1 \log Elec_{it-1} + \gamma_2 PopGrowth_{it} + u_i + \varepsilon_{it} \\ \text{où } y_{it} = \log(PIB_{it}/Pop_{it}) \quad t = 1998, 1999, \dots, 2013; \quad i = 1, 2, \dots, 61 \end{cases} \quad (5)$$

Les indices d'inégalités éducatives (Indice de Gini, Indice de Theil, Entropie généralisée d'ordre 0.5 {GE(0.5)} et l'indice d'Atkinson d'ordre 1 {A(1)}) (Ineq_Edu) ont été aussi estimés, comme pour le cas des régions et des communes, moyennant des données très désagrégés. Sur ce plan, plusieurs indices peuvent rendre compte de l'inégalité dans l'éducation. L'indice de Gini de l'éducation est celui le plus souvent utilisé dans la littérature. Toutefois, nous avons choisi, comme dans le chapitre II, les quatre indices d'inégalité pour tester la robustesse de nos résultats en répliquant les mêmes modèles non-paramétrique et semi-paramétrique et pour permettre des comparaisons par rapport aux résultats du chapitre II.

Outre *IneqEdu*, deux grandes variables explicatives sont considérées qui apparente l'équation retenue à une fonction de production dérivée en l'absence de séries d'emploi.

La consommation d'électricité par habitant, disponible annuellement au niveau provincial pour la période d'analyse⁶⁰, est utilisée comme proxy du capital physique et de son évolution. Elle a souvent été utilisée dans la littérature sur les déterminants de la croissance régionale et a été utilisée récemment dans les travaux s'intéressant à la relation entre inégalités éducatives et développement économique (Gille, 2015; Güngör, 2010). Le taux de croissance de la population a également été retenu pour chaque province à partir des données de la population du HCP. La variable *PopGrowth* est définie par le logarithme du taux de croissance démographique plus 0.05 {*Pop_Growth* = $\ln(\eta+0.05)$ } en suivant Mankiw et al. (1992).

3.4.3. Résultats du modèle

Le Tableau n° 29 rapporte l'estimation paramétrique linéaire à effets fixes sur le panel provincial. Les coefficients des indices d'inégalité d'éducation sont statistiquement significatifs et de signe négatif, ce qui indique qu'un niveau élevé d'inégalité éducative entraîne un faible développement économique. Les coefficients estimés pour le capital physique ainsi que l'accroissement de la population sont tous statistiquement significatifs et ont le signe attendu.

⁵⁹ Compte tenu des insuffisances du système statistique marocain en termes de données territorialisées, la DEPF utilise la méthode mixte à dominante descendante (répartition du PIB national) pour l'évaluation des PIB provinciaux. Plusieurs méthodes sont utilisées dans la littérature selon les spécificités des pays et de leurs systèmes d'information.

⁶⁰ Source : Ministère des Finances.

Tableau 29: Estimation paramétrique linéaire à effets fixes du panel provincial

Variables explicatives	Variable dépendante : Logarithme du PIB par habitant provincial			
	(1) ^a	(2) ^b	(3) ^c	(4) ^d
Constante	8.728 (30.99)***	8.064 (34.09)***	8.222 (34.28)***	6.416 (26.40)***
Ineq_Edu (t-1)	-3.109 (15.69)***	-1.202 (16.13)***	-0.963 (16.57)***	-3.828 (8.39)***
logElec (t-1)	0.140 (5.72)***	0.075 (2.73)***	0.089 (3.41)***	0.405 (24.25)***
Pop_Growth	-0.168 (5.39)***	-0.205 (6.76)***	-0.189 (6.25)***	-0.272 (8.26)***
Nb d'observations	915	915	915	915
Nb de provinces	61	61	61	61
R ² Within	0.66	0.67	0.67	0.60

Source : Auteur.

Les chiffres entre parenthèses correspondent aux statistiques de Student

***, ** et * indiquent que les variables sont respectivement significatives à 1%, 5 % et 10%

^aIneq_Edu correspond à l'indice de Gini de l'éducation

^bIneq_Edu correspond à l'indice de Theil de l'éducation

^cIneq_Edu correspond à l'entropie généralisée d'ordre 0.5 :GE(0.5)

^dIneq_Edu correspond à l'indice Atkinson d'ordre 1 : A(1)

La possibilité d'une éventuelle non-linéarité entre développement économique et inégalité de l'éducation est d'abord explorée en incluant les termes quadratiques et cubiques du niveau d'inégalité pour les quatre indices utilisés. Le modèle s'écrit sous la forme suivante :

$$\left\{ \begin{array}{l} y_{it} = \sum_{k=0}^{K+1} \beta_k (IneqEdu_{it-1})^k + \gamma_1 \log_cons_Elec_{it-1} + \gamma_2 PopGrowth_{it} + u_i + \varepsilon_{it} \\ \text{où } y_{it} = \ln(PIB_{it} / Pop_{it}) \quad t = 1998, 1999, \dots, 2013; \quad i = 1, 2, \dots, 61 \end{array} \right. \quad (6)$$

Les colonnes (1,4,7 et 10) du tableau n° 30 présentent les résultats de la spécification quadratique de l'équation (6.1). De même les colonnes (2,5,8 et 11) du tableau n° 30 présentent les résultats de la spécification cubique de l'équation (6.2). Les estimations des modèles ayant comme mesure de l'inégalité éducative, l'indice de Gini ou les indices d'entropie généralisée d'ordre 1 et 0.5, restent encore significatives pour la spécification cubique. Il faut signaler, après l'analyse des différents signes des termes linéaire, quadratique et cubique, qu'il y a une concordance entre les résultats de la spécification polynomiale trouvés pour les 1534 communes et ceux repérés sur les 61 provinces. En effet, dans ces deux modèles, les termes linéaire et cubique sont négatifs tandis que le terme quadratique est positif.

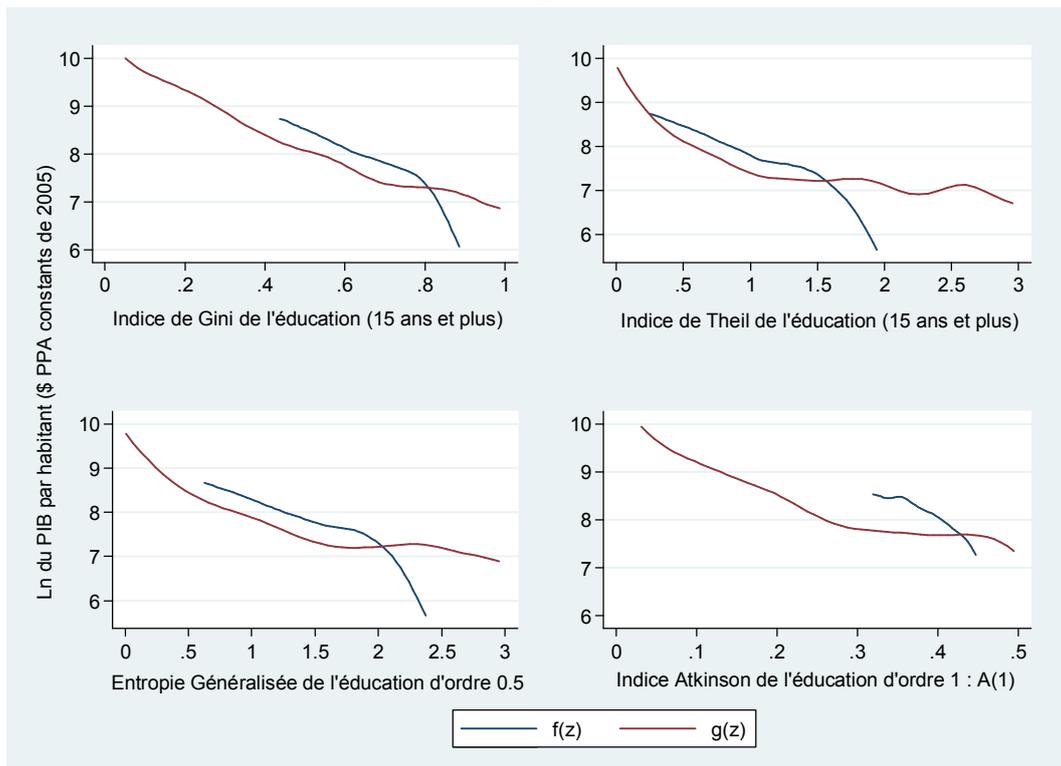
La fonction polynomiale dans l'équation (6) est maintenant remplacée dans le modèle par une fonction indéterminée à estimer $g(\cdot)$. Les modèles non-paramétrique et semi-paramétrique de notre relation s'écrivent respectivement selon les équations 7 et 8 suivantes:

$$\left\{ \begin{array}{l} y_{it} = g(IneqEdu_{it-1}) + u_i + \varepsilon_{it} \\ \text{où } y_{it} = \ln(PIB_{it} / Pop_{it}) \quad t = 1998, 1999, \dots, 2013; \quad i = 1, 2, \dots, 61 \end{array} \right. \quad (7)$$

$$\begin{cases} y_{it} = g(\text{IneqEdu}_{it-1}) + \gamma_1 \log \text{Elec}_{it-1} + \gamma_2 \text{PopGrowth}_{it} + u_i + \varepsilon_{it} \\ \text{où } y_{it} = \ln(\text{PIB}_{it} / \text{Pop}_{it}) \quad t = 1998, 1999, \dots, 2013; \quad i = 1, 2, \dots, 61 \end{cases} \quad (8)$$

Le graphique n° 27 présente les fonctions non-paramétriques $f(z)$ et $g(z)$ estimées à titre de comparaison pour le panel des 141 pays du monde sur la période 1955-2010 et le panel des 61 provinces marocaines sur la période 1998-2013. Rappelons que Berthélemy (2003) souligne que la trajectoire de la relation liant inégalités éducatives et développement ne peut pas être stable d'un pays à l'autre ni dans le temps.

Graphique 27: Fonctions non-paramétriques estimées pour le panel des pays du monde et celui des provinces marocaines



Source: Auteur.

$f(z)$: la fonction non-paramétrique estimée pour le panel des provinces marocaines

$g(z)$: la fonction non-paramétrique estimée pour le panel des pays du monde

Tableau 30: Estimations paramétrique et semi-paramétrique à effets fixes du panel provincial

Variable dépendante : Logarithme du PIB par habitant provincial

Variables explicatives	Modèle paramétrique		Modèle semi-paramétrique	Modèle paramétrique		Modèle semi-paramétrique	Modèle paramétrique		Modèle semi-paramétrique	Modèle paramétrique		Modèle semi-paramétrique
	Quadratique	Cubique		Quadratique	Cubique		Quadratique	Cubique		Quadratique	Cubique	
	(1) ^a	(2) ^a	(3) ^a	(4) ^b	(5) ^b	(6) ^b	(7) ^c	(8) ^c	(9) ^c	(10) ^d	(11) ^d	(12) ^d
constante	8.090 (25.79)***	11.358 (11.68)***	-	8.114 (33.93)***	8.343 (31.70)***	-	8.189 (33.94)***	8.528 (28.39)***	-	2.531 (2.96)***	4.932 (1.13)	-
Ineq_Edu (t-1)	0.370 (0.46)	-15.412 (3.41)***	-	-1.371 (9.66)***	-2.126 (5.43)***	-	-0.819 (6.19)***	-1.615 (3.66)***	-	18.342 (3.89)***	-2.003 (0.06)	-
Ineq_Edu ² (t-1)	-3.012 (4.43)***	22.536 (3.11)***	-	0.090 (1.40)	0.883 (2.27)**	-	-0.059 (1.22)	0.554 (1.69)*	-	-29.839 (4.73)***	26.722 (0.27)	-
Ineq_Edu ³ (t-1)	-	-13.560 (3.55)***	-	-	-0.255 (2.07)**	-	-	-0.149 (1.89)*	-	-	-51.790 (0.56)	-
log_cons_Elec (t-1)	0.082 (2.96)***	0.064 (2.29)**	0.076 (2.04)**	0.081 (2.93)***	0.079 (2.84)***	0.123 (3.33)***	0.078 (2.84)***	0.073 (2.64)***	0.087 (2.35)***	0.374 (21.12)***	0.374 (21.05)***	0.382 (17.37)***
Pop_Growth	-0.175 (5.68)***	-0.186 (6.06)***	-0.253 (6.74)***	-0.198 (6.43)***	-0.199 (6.48)***	-0.280 (7.40)***	-0.195 (6.36)***	-0.201 (6.52)***	-0.281 (7.44)***	-0.271 (8.35)***	-0.275 (8.31)***	-0.283 (7.03)***
Nb d'observations	915	915	915	915	915	915	915	915	915	915	915	915
Nb de provinces	61	61	61	61	61	61	61	61	61	61	61	61
R ² Within	0.67	0.67	-	0.67	0.67	-	0.67	0.67	-	0.61	0.61	-

Source: Auteur.

Les chiffres entre parenthèses correspondent aux statistiques de Student***, ** et * indiquent que les variables sont respectivement significatives à 1%, 5 % et 10%

^aIneq_Edu correspond à l'indice de Gini de l'éducation

^bIneq_Edu correspond à l'indice de Theil de l'éducation

^cIneq_Edu correspond à l'entropie généralisée d'ordre 0.5 :GE(0.5)

^dIneq_Edu correspond à l'indice Atkinson d'ordre 1 : A(1)

Puisque les fonctions non-paramétriques $f(z)$ et $g(z)$ estimés pour les deux panels sont représentées avec les mêmes échelles pour l'ensemble des variables (par exemple le log du PIB par habitant est en PPA \$ constant de 2005 à la fois pour le panel mondial que celui provincial), nous avons superposé ces deux fonctions en un seul graphique afin de faire apparaître les spécificités propres du cas marocain.

Les deux fonctions $f(z)$ et $g(z)$ reflètent globalement une tendance décroissante et non linéaire de la trajectoire. En effet, lorsque le niveau d'inégalité éducative est très élevé (cf. la partie droite des quatre courbes du graphique n° 27) la fonction a tendance à être concave. Celle-ci présente une concavité plus marquée pour les provinces marocaines par rapport à celle des pays du monde. Au fur et à mesure que l'inégalité baisse (cf. le sens est de droite à gauche pour les quatre courbes du graphique n° 27), sa concavité diminue jusqu'à atteindre un seuil d'inégalité à partir duquel la fonction tend à maintenir une tendance linéaire ayant la même pente (concavité nulle). En effet, les indices de Gini, Theil et l'entropie généralisée $GE(0.5)$ présentent des pentes presque identiques pour les deux panels. Ce constat est confirmé dans une moindre mesure pour l'indice d'Atkinson $A(1)$. Il semblerait ainsi que le deuxième seuil d'inégalité – identifié précédemment dans la modélisation au niveau communal – à partir duquel la fonction commence à devenir convexe n'est pas encore atteint pour ces trois indices alors que pour l'indice d'Atkinson $A(1)$ même le premier seuil n'est pas encore atteint.

Force est de constater que la fonction semi-paramétrique estimée pour notre panel provincial est, comme le cas du panel mondial du chapitre II, presque identique (à une translation près) à la fonction non-paramétrique. De même, les colonnes (3, 6, 9 et 12) du tableau n° 30 qui présentent les résultats de la partie paramétrique du modèle semi-paramétrique restent aussi significatifs et gardent les mêmes signes attendus. Nous nous contentons, dans ce qui suit, de présenter uniquement les résultats de la modélisation non-paramétrique puisqu'il s'est avéré que même en introduisant les deux variables de contrôle dont on dispose, l'allure de la fonction semi-paramétrique reste très semblable à celle de la fonction non-paramétrique.

3.4.4. Décomposition inter et intra-groupes des inégalités et contribution au développement économique des provinces marocaines

Afin de déceler le type d'inégalité éducative qui contribue le plus au développement des provinces marocaines, nous menons des décompositions⁶¹ *within* et *between* pour l'ensemble des 61 provinces à chacune des 16 années de la période d'analyse et ce pour les quatre indices d'inégalité. Par la suite, les deux composantes *within* et *between* sont intégrées dans la fonction à estimer en tant que couple de variables. Ainsi, la fonction non-paramétrique $g(IneqEdu_t)$ devient donc une surface non-paramétrique notée $g^S(IneqEdu_{it-1}^{Within}, IneqEdu_{it-1}^{Between})$ où $IneqEdu_{it-1} = IneqEdu_{it-1}^{Within} + IneqEdu_{it-1}^{Between}$.

Cette surface non-paramétrique est donc estimée à partir d'un ensemble de données qui fait correspondre à chaque couple d'inégalité d'éducation *within* et *between* $(IneqEdu_{it-1}^{Within}, IneqEdu_{it-1}^{Between})$ d'une province "i" à l'année "t", son niveau

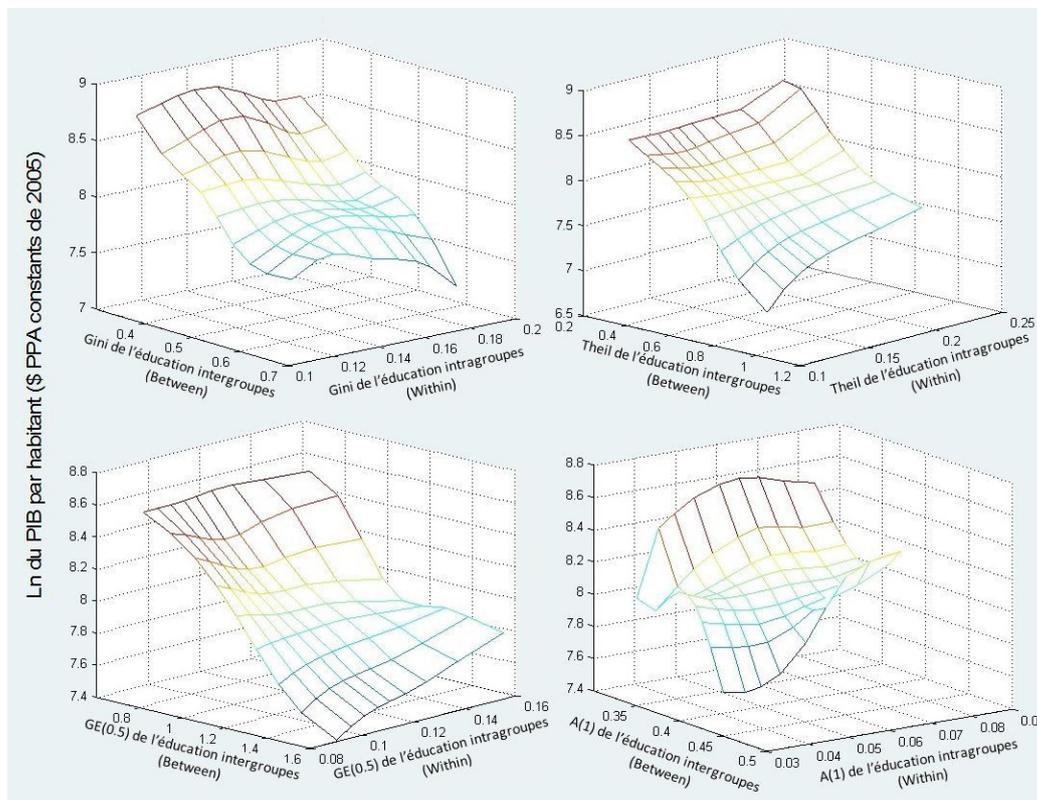
⁶¹ La méthodologie de la décomposition des indices d'inégalité éducative est décrite dans le chapitre I.

de développement économique y_{it} . Le modèle non-paramétrique de l'équation 7 devient donc :

$$\begin{cases} y_{it} = g^S(\text{IneqEdu}_{it-1}^{\text{Within}}, \text{IneqEdu}_{it-1}^{\text{Between}}) + u_i + \varepsilon_{it} \\ \text{où } y_{it} = \ln(\text{PIB}_{it}/\text{Pop}_{it}) \quad t = 1998, 1999, \dots, 2013; \quad i = 1, 2, \dots, 61 \end{cases} \quad (9)$$

Dans notre cas, il s'agit précisément de deux décompositions de l'inégalité d'éducation, la première concerne deux groupes (les illettrés et les non illettrés) et se décompose en une inégalité intra-éduqués et une inégalité inter-éduqués-illettrés et la seconde concerne 13 groupes (13 tranches d'âge allant de [15-19] à [75 et plus]) et se décompose en une inégalité intragénérationnelle et une inégalité intergénérationnelle.

Graphique 28 : Décomposition de l'effet des inégalités éducatives sur le développement des provinces: inégalités intra-éduqués et inter éduqués-illettrés



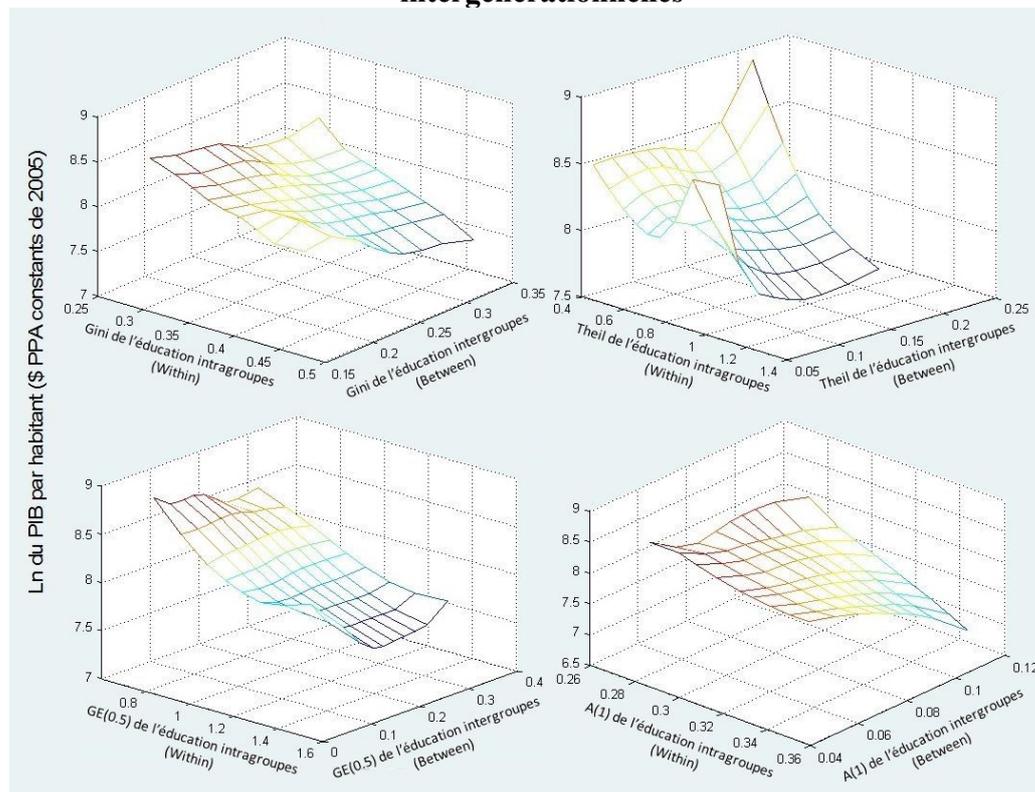
Source: Auteur.

Notons "I" l'indice d'inégalité de l'éducation qui peut être un indice de Gini, Theil, d'entropie généralisée ou d'Atkinson. Pour tracer la surface 3D non-paramétrique de l'indice d'inégalité "I" de l'une des deux décompositions, nous sommes basés sur le programme d'estimation sous Gauss de la fonction $\hat{g}^S(I)$, pour programmer l'estimation de 81 couples correspondant aux différentes combinaisons possibles des neuf déciles de la distribution de la composante *within* de l'indice I: $D_1^W(I)$ à $D_9^W(I)$ avec les neuf déciles de la distribution de la composante *between* de l'indice I: $D_1^B(I)$ à $D_9^B(I)$. Cette combinaison de couples

nous permet de tracer en 3D le triplet (z_1, z_2, y) qui fait correspondre à chaque couple de déciles $(z_1, z_2) = (D_i^W(I), D_j^B(I))$: la fonction $\hat{g}^S(D_i^W(I), D_j^B(I))$ tel que $i, j = 1, 2, \dots, 9$.

Les surfaces non-paramétriques tracées dans le graphique n° 28 représentent l'effet des inégalités éducatives intra-éduqués et inter éduqués-illettrés sur le développement provincial pour chacun des quatre indices d'inégalité. Ainsi, ces surfaces révèlent que le développement provincial est beaucoup plus expliqué par la réduction des inégalités *between* (inter éduqués-illettrés) que par la réduction des inégalités *within*. Ce constat peut être généralisé en faisant une décomposition pour *les moins éduqués / les plus éduqués* au lieu des *illettrés / éduqués*. On retrouve les mêmes conclusions. Ces dernières convergent pour affirmer que la promotion du développement économique passe d'abord par la réduction des inégalités *between* (entre les plus et les moins éduqués) que par une réduction des inégalités au sein de la même catégorie (au sein des plus éduqués ou au sein des moins éduqués).

Graphique 29: Décomposition de l'effet des inégalités éducatives sur le développement des provinces : inégalités intra-générationnelles et intergénérationnelles



Source : Auteur.

De plus, comme nous avons pu le démontrer dans le chapitre I, l'inégalité entre individus éduqués et illettrés dépend uniquement du taux d'analphabétisme et elle est une fonction croissante de ce taux (voir chapitre II, section 2.4.3). Par conséquent, il est plus bénéfique en termes de développement économique d'agir

sur la moitié inférieure de la distribution de l'éducation, notamment en visant à réduire le taux d'analphabétisme et les sorties précoces du primaire.

Ce constat établi pour le cas marocain ne correspond pas à celui de Castelló et Mukhopadhyay (2013) sur les Etats indiens. Pour ces auteurs le processus de développement dans le cas de l'Inde passe plutôt par la formation d'une élite que par une massification de l'éducation de base. Contrairement à nos résultats ces auteurs défendent l'idée qu'il faille agir davantage sur la moitié supérieure de la distribution de l'éducation pour stimuler le développement.

Nos résultats rejoignent par contre les conclusions de Berthélemy et Arestoff (2003) sur les stratégies d'éducation et de développement en Afrique. Ces auteurs critiquent les politiques éducatives élitistes en Afrique et recommandent que plus d'efforts soient consentis en priorité en faveur de l'éducation de base et que relativement moins de dépenses soient assignées à l'enseignement au lycée et au supérieur.

Une deuxième décomposition est menée pour 13 générations (de 15-19 ans à 75 et plus). Les surfaces non-paramétriques du graphique n° 29 représentent l'effet des inégalités éducatives intra-générationnelles et intergénérationnelles sur le développement provincial pour chacun des quatre indices d'inégalité. Il apparaît que, contrairement à la première décomposition, c'est plutôt la réduction des inégalités d'éducation à l'intérieur d'une même génération qui favorisent plus le développement.

3.5. Conclusion

Dans ce chapitre, la relation entre les inégalités éducatives et le développement a été abordée au niveau territorial. Notre analyse appliquée au Maroc a aussi bien considéré le niveau régional, provincial que communal.

Après une discussion des principaux problèmes méthodologiques qui entravent l'estimation de l'indice de Gini de l'éducation et de l'indice de développement humain au niveau infranational, nous avons montré que les inégalités éducatives au Maroc suivent un lent processus de convergence au niveau régional durant la période 1982-2010. Nous avons aussi mis en évidence qu'il existe une grande analogie entre le classement des régions marocaines sur la base du niveau de l'inégalité d'éducation et celui fondé sur l'IDH. Les inégalités éducatives s'avèrent être plus prononcées dans les régions les moins développées.

Nos résultats montrent aussi l'existence de fortes disparités de développement humain entre les régions marocaines. L'IDH régional varie de 0,473 pour la région de Taza Al Hoceima Taounate à 0,709 pour la région du Grand Casablanca qui occuperait, prise comme un pays, la 94^{ème} place mondiale. Cette région serait classée à la 30^{ème} position pour la santé, 99^{ème} pour le revenu par tête et 118^{ème} pour l'éducation.

Au-delà de l'analyse en termes de rang de classement, nous avons adopté une analyse en termes de nombre d'années de retard afin de mieux appréhender les

différences de l'IDH entre les régions. Certaines régions marocaines accusent un retard important dans les différentes composantes de développement humain par rapport à la région du Grand Casablanca. C'est le cas notamment de la région Taza Al Hoceima Taounate dont le retard atteint 21 ans pour l'espérance de vie à la naissance, 41 ans pour la moyenne d'années de scolarisation, 17 ans pour l'espérance de vie scolaire et de 47 ans pour le revenu. De manière générale, par rapport au Grand Casablanca, le retard pour la majorité des régions est plus prononcé concernant la moyenne d'années de scolarisation.

Par ailleurs, à travers une analyse exploratoire des données spatiales effectuée sur une coupe transversale de 1534 communes, il a été montré, que le développement humain et les inégalités éducatives sont très localisés sur le territoire marocain et inversement corrélés. De plus, en utilisant un modèle économétrique spatial, nos résultats montrent que la réduction des inégalités d'éducation impacte sensiblement le niveau de développement humain. Nous avons aussi mis en évidence que la relation affiche une tendance non linéaire où l'impact est moins élevé pour les communes relativement inégalitaires.

Ce chapitre s'est aussi intéressé à la relation entre les inégalités éducatives et le développement économique estimée sur un panel de 61 provinces marocaines durant la période 1998-2013. Nous avons exploré l'existence d'un éventuel comportement non linéaire de la relation par le biais d'une modélisation non-paramétrique et semi-paramétrique en données de panel. De manière générale, nous avons mis en évidence qu'une relation inverse est confirmée et que plus la province est inégalitaire en termes d'accès à l'éducation, plus elle est susceptible d'enregistrer de faibles performances économiques. De plus, cet impact se révèle suivre un changement de concavité dans la trajectoire de la relation. En effet, la réduction de l'inégalité d'éducation est plus bénéfique pour le développement économique des provinces lorsque le niveau initial est très inégalitaire ($Gini > 0.8$; $Theil > 1.5$; $GE(0.5) > 2$) et le niveau de PIB par habitant inférieur à 2100 \$ PPA constants de 2005.

Les résultats des surfaces non-paramétriques suggèrent des pistes d'action pour les politiques publiques marocaines. Pour favoriser le développement économique, il s'agit en priorité de lutter contre l'analphabétisme dans les générations relativement jeunes et contre les sorties précoces du système éducatif.



Conclusion générale

Conclusion générale

Cette thèse a cherché à explorer le lien entre inégalités éducatives et développement économique qui a fait l'objet d'analyses controversées en raison notamment de l'indisponibilité de données transnationales comparables. Ce lien a la particularité de recouper en même temps deux relations du champ de l'économie du développement : celle dérivée de la relation éducation/croissance lorsque l'inégalité dans l'éducation est considérée comme une mesure non quantitative de « l'éducation », et celle dérivée de la relation inégalité/croissance lorsque celle-ci est retenue comme variable non monétaire de « l'inégalité ».

Dans le premier chapitre, nous avons proposé une nouvelle base de données comportant plusieurs mesures des inégalités éducatives (indice de Gini, de Theil, entropie généralisée d'ordre 0.5 et indice d'Atkinson d'ordre 1) pour un large panel de pays sur la période 1950-2010. Nous avons cherché à améliorer sensiblement les estimations des indices d'inégalité d'éducation et de la moyenne d'années de la scolarisation. Nous avons montré que la pondération utilisée par Barro et Lee (2013), pour obtenir les durées cumulées de scolarisation d'une tranche d'âge agrégée, conduit à un biais dans l'estimation de la moyenne d'années de scolarisation. Ce biais peut même atteindre une différence de l'ordre d'une demi-année en termes de moyenne d'années de scolarisation. Notre base de données corrige cette erreur de pondération et relâche certaines hypothèses peu vraisemblables adoptées par plusieurs auteurs. Nous avons ainsi pris en compte la variation au cours du temps et entre les différentes classes d'âge des durées de scolarisation et l'hétérogénéité par rapport aux niveaux d'éducation au sein d'un même cycle d'enseignement.

Par ailleurs, notre base de données met aussi à la disposition des chercheurs des estimations pour les différentes tranches d'âge ([15-19]; [20-24]; [25-29] ans..) Ainsi, à titre d'exemple, Kochav et al. 2015 ont introduit, dans leurs modélisations économétriques, l'indice de Gini de l'éducation de la tranche d'âge [25-29] ans issu de notre base de données⁶². Certains auteurs ont utilisé nos estimations de l'inégalité d'éducation comme principale variable explicative (Bartusevičius, 2014; Bosancianu, 2017) ou comme une variable de contrôle parmi d'autres (Holton, 2012; Kochav et al. 2015; Oyvatt, 2016).

A partir de cette nouvelle base de données, nous avons également mis en évidence la dynamique d'évolution des inégalités d'éducation au niveau mondial selon le niveau de développement et les groupes d'âge quinquennaux. Ainsi, contrairement aux inégalités de revenu qui affichent des évolutions tantôt à la hausse tantôt à la baisse (Morrisson et Murin, 2013), nous avons montré que les

⁶² Signalons que notre base de données a déjà été téléchargée plus de 1000 fois et citée dans des rapports et articles scientifiques plus de trente fois depuis 2012 (Nations Unies, 2013; Eberstadt, 2017; Trabelsi, 2013; Peterson, 2014; Potter, 2014; Cruces et al., 2014; Banque asiatique de développement, 2014; Marsh, 2016; Bernauer, 2016; Ascher et Mirovitskaya, 2016; Garces-Voisinat, 2016; Jorda et Alonso, 2017; Martins, 2016; Sauer, 2016).

inégalités d'éducation sont marquées par une baisse conséquente et continue pour la quasi-totalité des pays du monde et tout au long des six dernières décennies. Nous avons aussi mis en évidence que l'écart entre pays en développement et pays développés, en termes de niveau d'inégalité éducative est moins prononcé pour la classe d'âge des 15-24 ans où l'on décèle une relative convergence. Cette convergence s'avère être moins perceptible pour d'autres tranches d'âge.

Suite à la construction de cette base de données, il a été question de savoir si la réduction de l'inégalité d'éducation a un impact sur le développement économique des PVD. Avant d'entamer l'analyse de ce lien, nous avons discuté certains choix méthodologiques. La plupart des travaux sur le lien entre les inégalités éducatives et le développement économique introduisent dans le même modèle économétrique avec la moyenne d'années de scolarisation l'indice de Gini de l'éducation. Cette spécification fragilise les résultats de l'estimation dans la mesure où les deux variables sont souvent fortement corrélées (Berthélemy, 2006). Afin de contourner ce problème de corrélation, Gille (2015) a proposé d'intégrer simultanément, dans le modèle, la moyenne d'années de scolarisation avec un autre indice, proposé par Berthélemy (2006) qui est dérivé du Gini de l'éducation mais apuré de « l'effet du niveau moyen ». Toutefois, nous avons mis en évidence que même cet indice demeure corrélé à la moyenne d'années de scolarisation. Ainsi nous avons choisi de retenir uniquement l'indice d'inégalité de l'éducation en plus des variables de contrôle. Ce choix méthodologique nous a permis, à la fois, de contourner le problème de corrélation et d'explorer spécifiquement une dimension très peu abordée dans les modèles de croissance.

Une autre option méthodologique a été également au cœur de toutes les mesures d'inégalité éducative adoptées dans cette thèse. Nous avons opté pour quatre indices d'inégalité : l'indice de Gini, de Theil, l'entropie généralisée d'ordre 0.5 et l'indice d'Atkinson de l'éducation d'ordre 1. Même si ce dernier indice présente certaines spécificités par son mode de calcul, nous l'avons tout de même maintenu étant donné qu'il est de plus en plus considéré dans les approches statistiques ces dernières années⁶³.

Dans le chapitre II, après avoir passé en revue les différents travaux empiriques sur la relation entre les inégalités d'éducation et la croissance économique, nous avons mis essentiellement l'accent sur l'analyse de la non linéarité de la relation. Dans une première étape, l'estimation du modèle cubique nous a révélé une prédisposition de la relation à avoir une allure qui change de concavité puisque les puissances d'ordre deux et trois des indices de Gini et d'entropie généralisée d'ordre 1 et 0.5 apparaissent statistiquement significatives.

Dans une deuxième étape, la modélisation non-paramétrique nous a permis de mieux distinguer les changements de trajectoire de la relation et d'identifier trois principales phases qui dépendent essentiellement du niveau d'inégalité d'éducation. Néanmoins étant donné que les pays en développement sont relativement

⁶³ L'indice Atkinson d'ordre 1 a été retenu par le PNUD durant les sept dernières années pour le calcul de l'IDH ajusté aux inégalités.

hétérogènes, cette tendance demeure relativement globale et peut ne pas refléter les spécificités de certaines zones géographiques.

Dans une troisième étape, les fonctions non-paramétriques estimées par régions du monde nous ont permis de décortiquer les phases que parcourt la relation. Un schéma récapitulatif des différentes phases a été introduit. En se basant sur les intervalles de confiance qui permettent de repérer les zones les plus stables des trajectoires régionales, nous avons retenu au total six phases.

Les trois premières phases sont repérées par rapport au niveau de développement et concernent essentiellement les pays dont le niveau du PIB par habitant (\$ PPA constants de 2005) appartient respectivement aux intervalles [0, 1100], [1100, 1500] et [1500, 3000]. La réduction d'inégalité éducative dans la première phase a peu d'effet sur le développement économique. Cette phase concerne principalement les pays de l'Afrique subsaharienne dont le niveau de PIB par habitant reste en deçà d'un seuil de développement (1100 \$ PPA constants de 2005) à partir duquel tend à se déclencher un processus perceptible de décollage économique. La réduction de l'inégalité éducative dans la deuxième phase apparaît avoir un effet modéré sur le développement économique. Durant la troisième phase, à partir du seuil de 1500 \$ PPA constants de 2005, l'impact de la réduction de l'inégalité sur le PIB par tête est fortement croissant. Même si dans cette phase, la distribution de l'éducation reste relativement inégalitaire.

La quatrième phase, tout comme la première, dévoile un effet pratiquement inexistant de la diminution de l'inégalité d'éducation sur la croissance économique. Celle-ci concerne les pays qui ont un PIB par habitant appartenant à l'intervalle [3000, 8000] et un indice de Gini de l'éducation dans l'intervalle [0.4, 1]. Cette phase concerne plus particulièrement les pays de la zone MENA et certains pays d'Amérique latine. En revanche, la cinquième phase ressemble plus à la troisième. Elle concerne principalement les pays dont le PIB par habitant est supérieur au seuil critique de développement de 8000 \$ PPA (constants de 2005) et dont le niveau d'inégalité d'éducation est relativement faible (Gini de l'éducation appartenant à l'intervalle [0.2, 0.4]).

La dernière phase, identifiée par une distribution très équitable de l'éducation, présente un impact ambigu de l'inégalité d'éducation sur la croissance. Cette phase est principalement déterminée par le niveau d'inégalité éducative (Gini de l'éducation inférieur à 0.2) et concerne majoritairement les pays développés et quelques pays de l'Asie centrale et d'Europe moins développés.

Nous avons donc mis en évidence que c'est dans les phases 3 et 5 que l'effet de la réduction des inégalités éducatives est le plus bénéfique sur le développement économique. C'est le cas où la distribution de l'éducation est très inégalitaire avec un niveau du PIB par habitant supérieur à 1500 \$ PPA, où lorsqu'elle est relativement égalitaire avec un PIB par habitant dépassant 8000 \$ PPA. Pour les pays se situant dans les phases 1, 2, 4 et 6, plusieurs autres facteurs que celui de la distribution inégalitaire de l'éducation peuvent entrer en ligne de compte.

Suite à la décomposition de l'inégalité éducative selon le niveau d'éducation et selon les tranches d'âge, nous avons pu discuter certaines orientations de

politique publique et quelques questionnements sur les arbitrages éducatifs liés à la distribution de l'éducation. Nous avons ainsi mis en évidence que la réduction de l'inégalité entre les plus éduqués et les moins éduqués (*between*) est plus profitable aux économies en développement qu'une réduction de l'inégalité au sein des plus éduqués ou au sein des moins éduqués (*within*). Nous avons aussi montré que cela revient à agir sur la moitié inférieure de la distribution de l'éducation en augmentant le niveau d'alphabétisation. Malgré les efforts consentis pour assurer l'accès égal à l'alphabétisation et à l'éducation de base, les pays en développement restent en deçà des objectifs fixés⁶⁴. En effet, le quatrième objectif du millénaire pour le développement, qui a visé de réduire de moitié le taux d'analphabétisme entre 2000 et 2015, n'a pas été atteint par plus des trois quarts des pays du monde.

Par ailleurs, l'estimation de la surface non-paramétrique pour la décomposition par rapport aux tranches d'âges ([15-19], [20-24], .. [75 et plus]) révèle plutôt que c'est la diminution de l'inégalité *within* qui est plus bénéfique à la croissance économique. Les politiques éducatives ont plus tendance à s'attacher aux différences entre les générations⁶⁵ qu'aux comparaisons intra-génération. Nos résultats dévoilent que les efforts devraient plutôt être consentis à la réduction des inégalités éducatives au sein d'une même génération, et surtout lorsque celle-ci présente une proportion élevée d'analphabètes.

Etant donné que le développement économique n'est pas une fin en soi et qu'il n'entraîne pas *de facto* un meilleur développement humain pour les individus et les nations, nous nous sommes aussi intéressés, dans le chapitre III, aux autres dimensions non monétaires du développement. En effet, le chapitre III a été dédié à l'étude de la relation entre les inégalités d'éducation et le développement local dans le cas du Maroc où plusieurs analyses infranationales ont été développées selon trois niveaux de territorialité (régional, provincial et communal).

Notre analyse au niveau régional a porté essentiellement sur l'évaluation à la fois du niveau d'inégalité d'éducation et du développement humain à travers les indices de Gini de l'éducation et du développement humain. Nous avons mis en évidence que les régions marocaines accusent un retard important dans la composante « éducation » du développement humain. De même, nous avons montré que les inégalités éducatives sont plus accentuées dans les régions les moins développées. Par ailleurs, l'analyse régionale menée en termes de nombre d'années de retard nous a révélé que c'est particulièrement dans la sous-dimension « moyenne d'années de scolarisation » de l'IDH que le retard est plus prononcé. Il peut même atteindre une quarantaine d'années.

Dans le chapitre III, nous avons aussi analysé au niveau provincial la relation entre les inégalités éducatives et le développement économique. En effet, en adoptant les mêmes étapes utilisées dans le chapitre II, nos résultats ont mis en évidence l'existence d'un comportement non linéaire de cette relation en

⁶⁴ Le 50^{ème} anniversaire de la Journée Internationale de l'Alphabétisation célébré en septembre 2016, a dressé un bilan critique sur les avancées accomplies par les pays du monde dans la lutte contre l'analphabétisme des adultes. Le taux d'analphabétisme de la population mondiale est évalué à l'ordre de 15% même si au début des années cinquante, près de la moitié de cette même population ne savait ni lire ni écrire.

⁶⁵ A titre d'exemple, l'évolution du score/classement PISA au fil du temps revient en quelque sorte à une comparaison entre les différentes générations.

considérant les provinces marocaines. Celles-ci s'avèrent être principalement situées à la phase 3 identifiée au chapitre II. De plus, nous avons constaté que lorsque le niveau du PIB par habitant est inférieur à 2100 \$⁶⁶, la diminution du niveau d'inégalité éducative est plus favorable pour le développement économique. Ainsi, l'on pourrait constater que la phase 3 identifiée au chapitre II est scindée, dans le cas des provinces marocaines, en deux sous phases par rapport au seuil de 2100 \$. Par ailleurs, les résultats des surfaces non-paramétriques obtenus au chapitre II sont identiques à ceux du chapitre III. La lutte contre l'analphabétisme et les sorties précoces du système éducatif devrait donc être une priorité pour les politiques éducatives au Maroc.

Une analyse plus fine au niveau communal a été menée à travers une analyse exploratoire des données spatiales (la méthode ESDA) et par la spécification d'un modèle économétrique spatial dans un second temps. Celle-ci a révélé que les niveaux d'inégalité éducative et du développement humain sont réciproquement reliés l'un à l'autre et très localisés sur le territoire marocain.

Bien que les analyses effectuées dans cette thèse diffèrent selon le niveau géographique (mondial / régional / provincial / communal), selon la variable dépendante (développement économique / développement humain), selon le modèle économétrique (non-paramétrique / polynomial spatial) et selon le type de représentation des données (en coupe transversale / en panel), nos résultats nous amènent à souligner deux principaux constats communs. D'une part, la trajectoire de la relation entre les inégalités éducatives et le développement est généralement en sens inverse sous certaines conditions soulignées tout au long de la thèse. D'autre part, elle est de plus caractérisée par une tendance non linéaire. Néanmoins, certains aspects mériteraient d'être explorés plus avant, et pourraient être menés dans le cadre de futures extensions. Nous en exposons brièvement ici quelques-unes.

Une première extension possible serait d'étudier, pour chacune des six phases identifiées de la relation, les différents mécanismes à travers lesquels l'inégalité éducative influence le développement économique. A cet égard, il serait particulièrement intéressant de comparer les mécanismes démographiques avec celui des imperfections des marchés du crédit, et d'identifier les phases dans lesquelles chacun de ces deux mécanismes intervient.

Comme deuxième extension, il serait opportun de mener une analyse de la non linéarité de la relation sur la base de données sur les inégalités de performances scolaires. Etant donné la difficulté dans la disponibilité de telles données au niveau transnational, l'étude de cas d'un pays particulier comme le Maroc serait aussi intéressante. Dans cette perspective, on pourrait par la suite déceler les phases que parcourt cette relation en adoptant les mêmes démarches.

Les premières données sur le nouveau découpage du Maroc viennent d'être publiées récemment. Elles proviennent essentiellement du recensement de 2014 de la population et de l'habitat. En prenant en considération ces données, l'analyse

⁶⁶ La mesure adoptée est la même : PPA constants de 2005.

menée au niveau communal au chapitre III pourrait ainsi être singulièrement prolongée.



Annexes

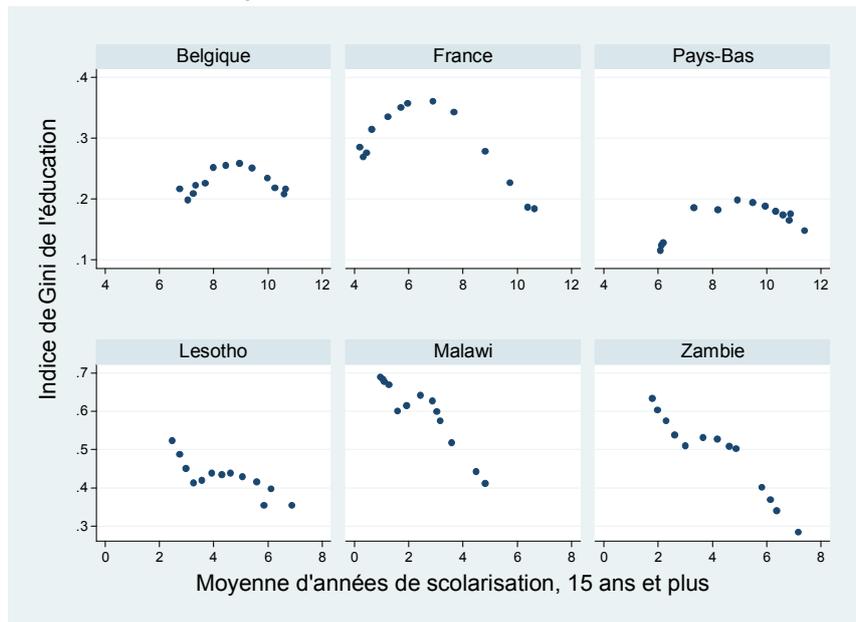
Annexes

Annexe 1: Formules des indices d'inégalité synthétisées selon les différents cas de figure

Type	L'indice de Gini de l'éducation	L'entropie généralisée de l'éducation	Indice Atkinson de l'éducation
Groupe d'âge quinquennal d'un pays	$Gini_{c,t}^a = \frac{\sum_{i=1}^n \sum_{j=1}^n n_{c,i,t}^a y_{c,i,t}^a - y_{c,j,t}^a n_{c,j,t}^a}{2 N_{c,t}^a \sum_{i=1}^n n_{c,i,t}^a y_{c,i,t}^a}$	$GE(r)_{c,t}^a = \frac{N_{c,t}^{a(r-1)}}{r(r-1)} \frac{\left[\sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^r \right]}{\left(\sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a \right)^r} - 1$	$A(\varepsilon)_{c,t}^a = 1 - \frac{1}{\left[N_{c,t}^a \right]^{\left(\frac{\varepsilon}{1-\varepsilon} \right)}} \frac{\left[\sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^{(1-\varepsilon)} \right]^{\left(\frac{1}{1-\varepsilon} \right)}}{\sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a}$
Grand Groupe d'âge d'un pays	$Gini_{c,t}^G = \frac{\sum_{a \in G} \sum_{a' \in G} \sum_{i=1}^n \sum_{j=1}^n n_{c,i,t}^a y_{c,i,t}^a - y_{c',j,t}^{a'} n_{c',j,t}^{a'}}{2 N_{c,t}^G \left[\sum_{a \in G} \sum_{i=1}^n n_{c,i,t}^a y_{c,i,t}^a \right]}$	$GE(r)_{c,t}^G = \frac{N_{c,t}^{G(r-1)}}{r(r-1)} \frac{\left[\sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^r \right]}{\left(\sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a \right)^r} - 1$	$A(\varepsilon)_{c,t}^G = 1 - \frac{1}{\left[N_{c,t}^G \right]^{\left(\frac{\varepsilon}{1-\varepsilon} \right)}} \frac{\left[\sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^{(1-\varepsilon)} \right]^{\left(\frac{1}{1-\varepsilon} \right)}}{\sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a}$
Groupe d'âge quinquennal d'une zone géographique	$Gini_{R,t}^a = \frac{\sum_{c \in R} \sum_{c' \in R} \sum_{i=1}^n \sum_{j=1}^n n_{c,i,t}^a y_{c,i,t}^a - y_{c',j,t}^a n_{c',j,t}^a}{2 N_{R,t}^a \left[\sum_{c \in R} \sum_{i=1}^n n_{c,i,t}^a y_{c,i,t}^a \right]}$	$GE(r)_{R,t}^a = \frac{N_{R,t}^{a(r-1)}}{r(r-1)} \frac{\left[\sum_{c \in R} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^r \right]}{\left(\sum_{c \in R} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a \right)^r} - 1$	$A(\varepsilon)_{R,t}^a = 1 - \frac{1}{\left[N_{R,t}^a \right]^{\left(\frac{\varepsilon}{1-\varepsilon} \right)}} \frac{\left[\sum_{c \in R} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^{(1-\varepsilon)} \right]^{\left(\frac{1}{1-\varepsilon} \right)}}{\sum_{c \in R} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a}$
Grand Groupe d'âge d'une zone géographique	$Gini_{R,t}^G = \frac{\sum_{c \in R} \sum_{c' \in R} \sum_{a \in G} \sum_{a' \in G} \sum_{i=1}^n \sum_{j=1}^n n_{c,i,t}^a y_{c,i,t}^a - y_{c',j,t}^{a'} n_{c',j,t}^{a'}}{2 N_{R,t}^G \left[\sum_{c \in R} \sum_{a \in G} \sum_{i=1}^n n_{c,i,t}^a y_{c,i,t}^a \right]}$	$GE(r)_{c,t}^G = \frac{N_{R,t}^{G(r-1)}}{r(r-1)} \frac{\left[\sum_{c \in R} \sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^r \right]}{\left(\sum_{c \in R} \sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a \right)^r} - 1$	$A(\varepsilon)_{R,t}^G = 1 - \frac{1}{\left[N_{R,t}^G \right]^{\left(\frac{\varepsilon}{1-\varepsilon} \right)}} \frac{\left[\sum_{c \in R} \sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a (y_{c,j,t}^a)^{(1-\varepsilon)} \right]^{\left(\frac{1}{1-\varepsilon} \right)}}{\sum_{c \in R} \sum_{a \in G} \sum_{j=1}^n n_{c,j,t}^a y_{c,j,t}^a}$

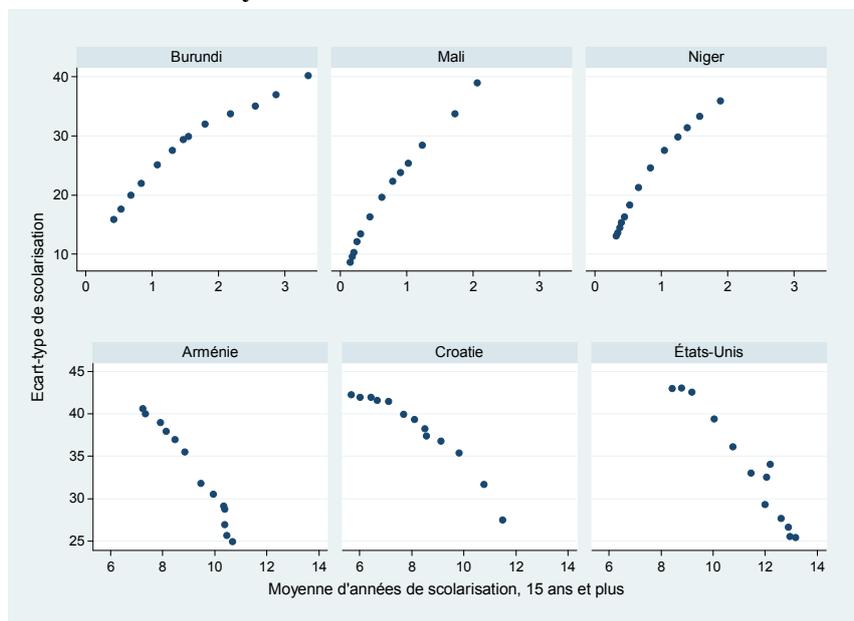
Source : Auteur.

Annexe 2.1 : Cas atypiques du lien entre l'indice de Gini de l'éducation et la moyenne d'années de scolarisation



Source: Auteur (Base mondiale des inégalités éducatives).

Annexe 2.2 : Cas atypiques du lien entre l'écart-type de scolarisation et la moyenne d'années de scolarisation



Source: Auteur (Base mondiale des inégalités éducatives).

Annexe 3 : Démonstration de la formulation de l'indice de Gini de l'éducation

Soit $s_k = \begin{cases} 1 & \text{si } k < i \\ (\mu - y_i)/(y_n w_i) & \text{si } k = i \\ 0 & \text{si } k > i \end{cases}$, tel que la moyenne d'années de scolarisation μ vérifie $y_i \leq \mu < y_{i+1}$. On peut démontrer que dans ce cas l'indice de Gini de l'éducation est égal à $= \frac{(\mu - y_i)(y_{i+1} - \mu)}{\mu(y_{i+1} - y_i)}$.

On calcule d'abord $y_n \sum_{k=1}^{n-1} w_k s_k$ et $y_n \sum_{k=1}^{n-1} w_k s_k^2$

$$y_n \sum_{k=1}^{n-1} w_k s_k = y_n \left[\sum_{k=1}^{i-1} w_k \right] + y_n w_i s_i = y_n \frac{y_i}{y_n} + (\mu - y_i) = \mu$$

$$y_n \sum_{k=1}^{n-1} w_k s_k^2 = y_n \left[\sum_{k=1}^{i-1} w_k \right] + y_n w_i s_i^2 = y_n \frac{y_i}{y_n} + y_n w_i \left[\frac{\mu - y_i}{y_n w_i} \right]^2 = y_i + \frac{(\mu - y_i)^2}{(y_{i+1} - y_i)}$$

Ainsi on peut calculer l'indice de Gini de l'éducation comme suit :

$$Gini = 1 - \frac{\sum_{k=1}^{n-1} w_k s_k^2}{\sum_{k=1}^{n-1} w_k s_k} = \frac{\mu - \left[y_i + \frac{(\mu - y_i)^2}{(y_{i+1} - y_i)} \right]}{\mu}$$

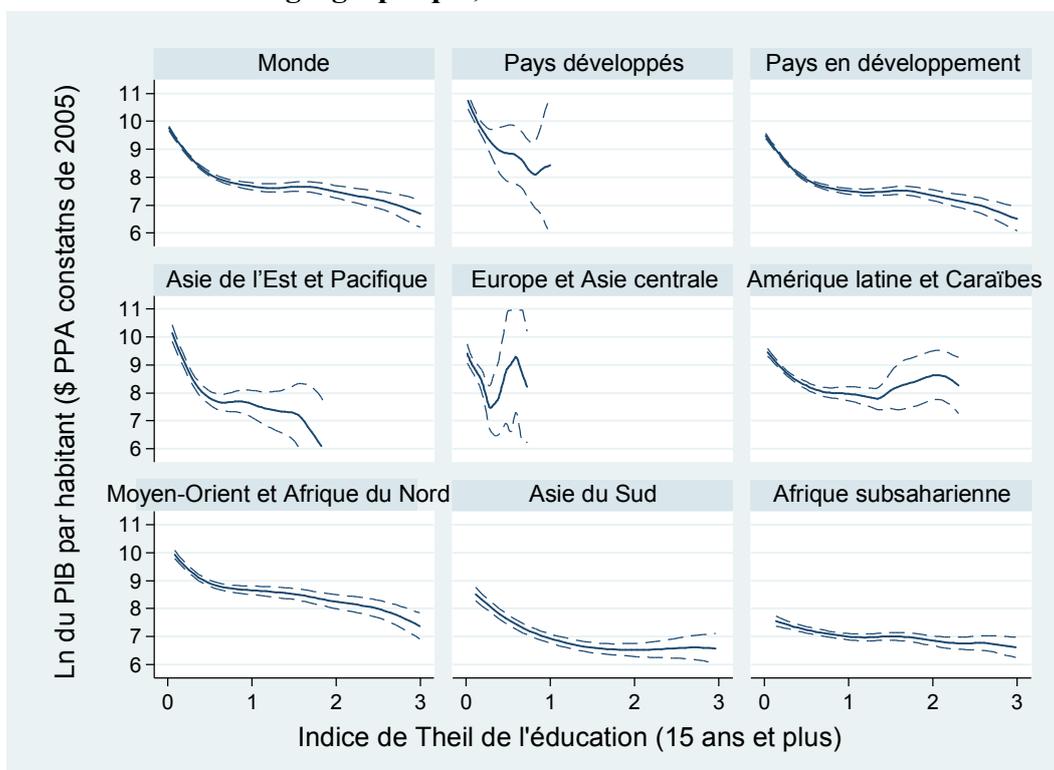
$$Gini = \frac{\mu(y_{i+1} - y_i) - (y_{i+1} - y_i) \left[y_i + \frac{(\mu - y_i)^2}{(y_{i+1} - y_i)} \right]}{\mu(y_{i+1} - y_i)}$$

$$Gini = \frac{\mu(y_{i+1} - y_i) - y_i(y_{i+1} - y_i) - \mu^2 + 2\mu y_i - y_i^2}{\mu(y_{i+1} - y_i)} = \frac{\mu y_{i+1} + \mu y_i - y_i y_{i+1} - \mu^2}{\mu(y_{i+1} - y_i)}$$

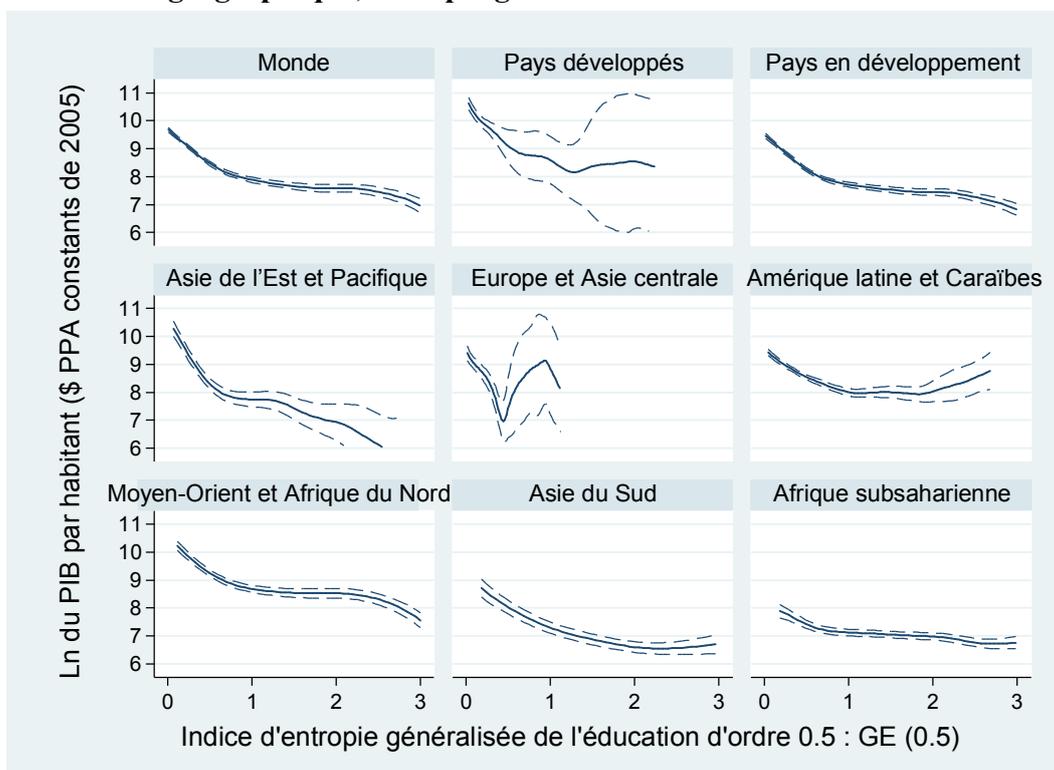
d'où

$$Gini = \frac{(\mu - y_i)(y_{i+1} - \mu)}{\mu(y_{i+1} - y_i)}$$

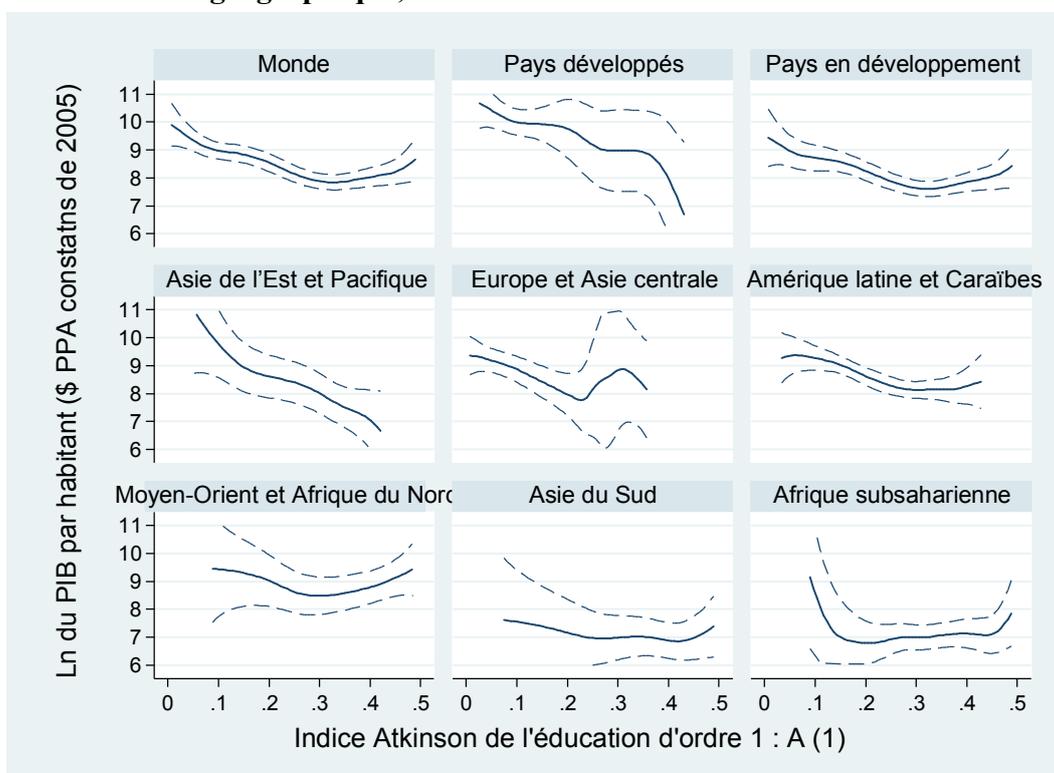
Annexe 4.1 : Fonction non-paramétrique selon le niveau de développement et la zone géographique, indice de Theil de l'éducation



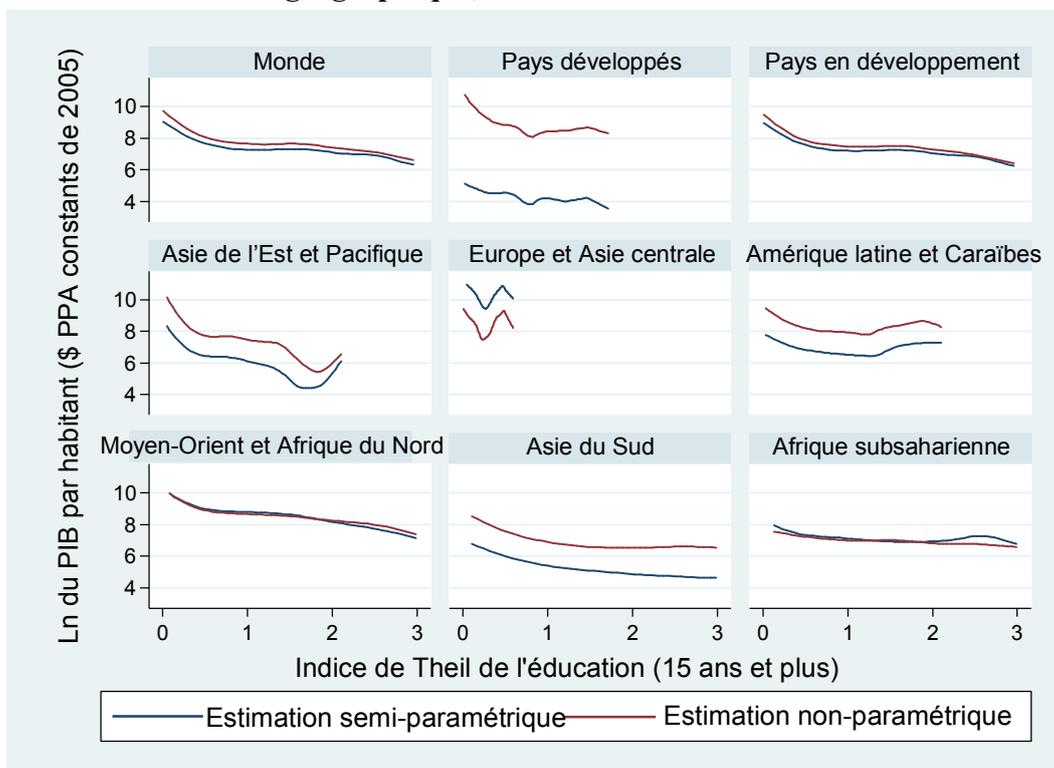
Annexe 4.2 : Fonction non-paramétrique selon le niveau de développement et la zone géographique, entropie généralisée de l'éducation d'ordre 0.5



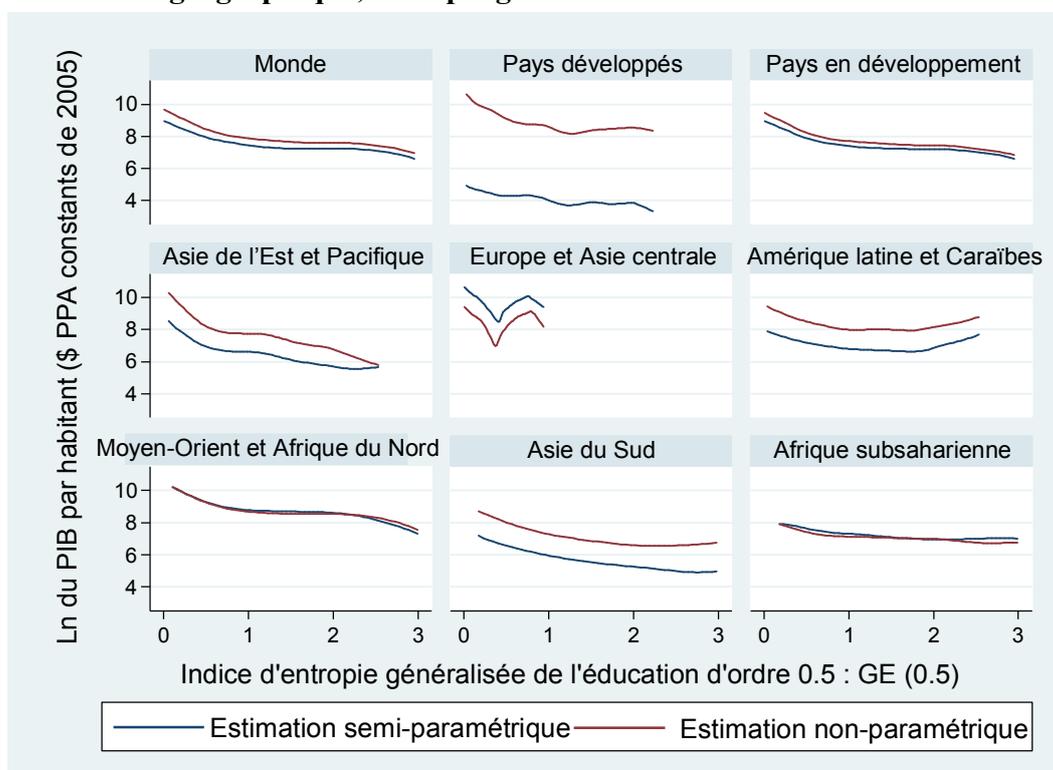
Annexe 4.3 : Fonction non-paramétrique selon le niveau de développement et la zone géographique, indice Atkinson de l'éducation d'ordre 1



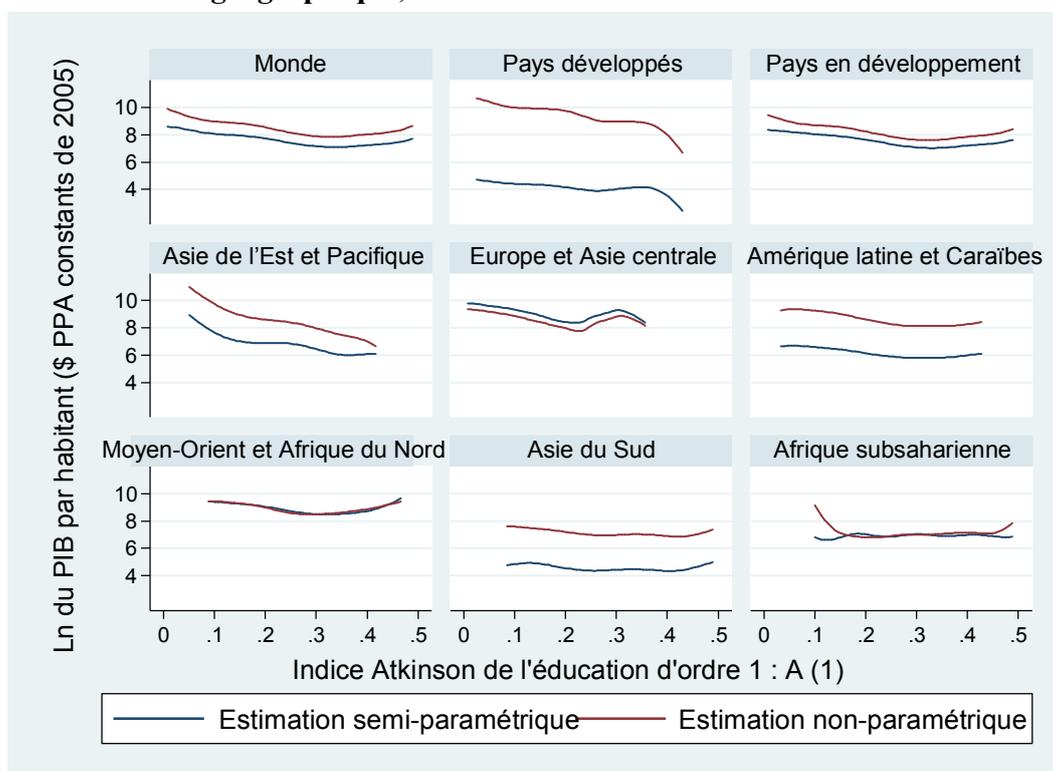
Annexe 5.1 : Fonction semi-paramétrique selon le niveau de développement et la zone géographique, indice de Theil de l'éducation



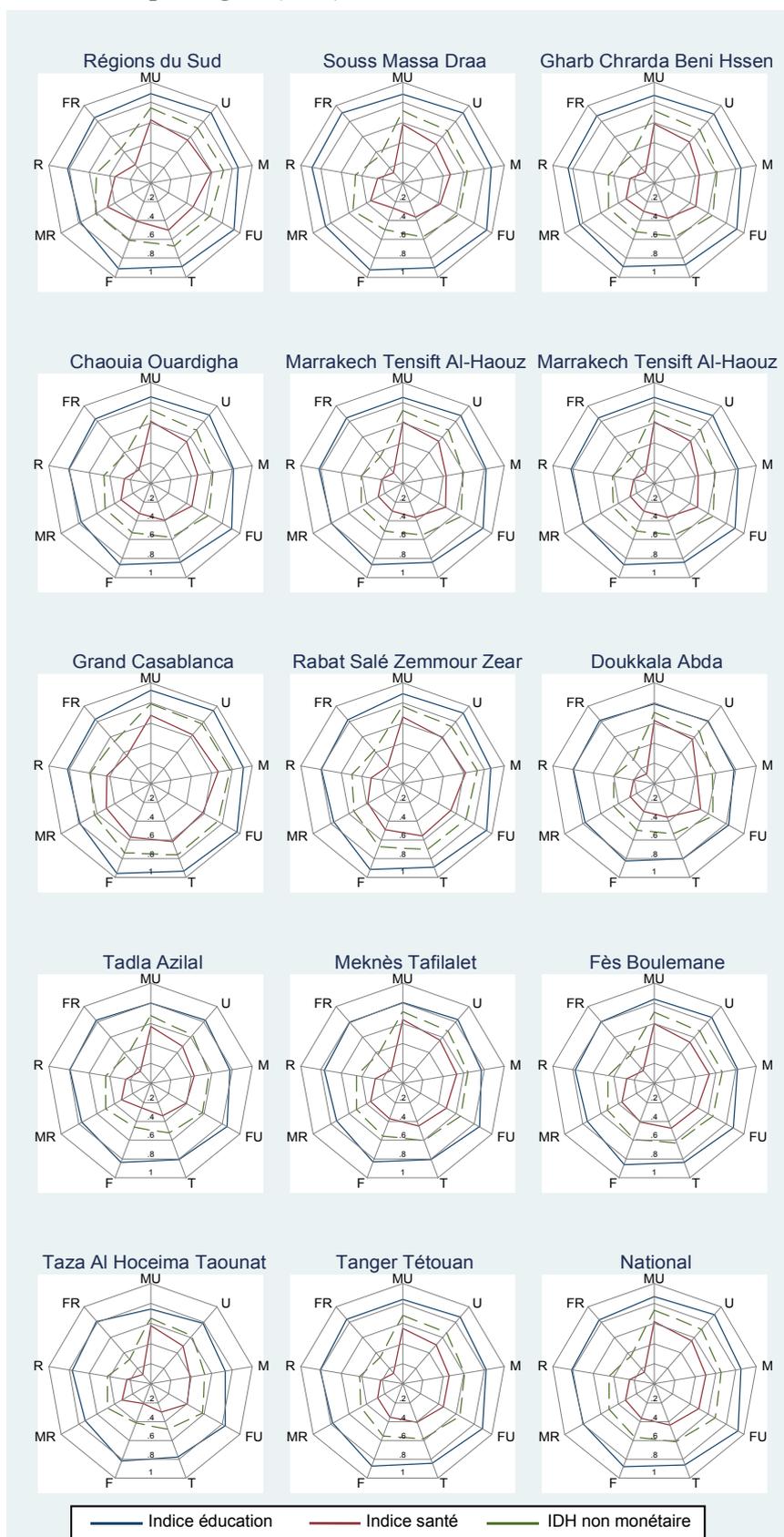
Annexe 5.2 : Fonction non-paramétrique selon le niveau de développement et la zone géographique, entropie généralisée de l'éducation d'ordre 0.5



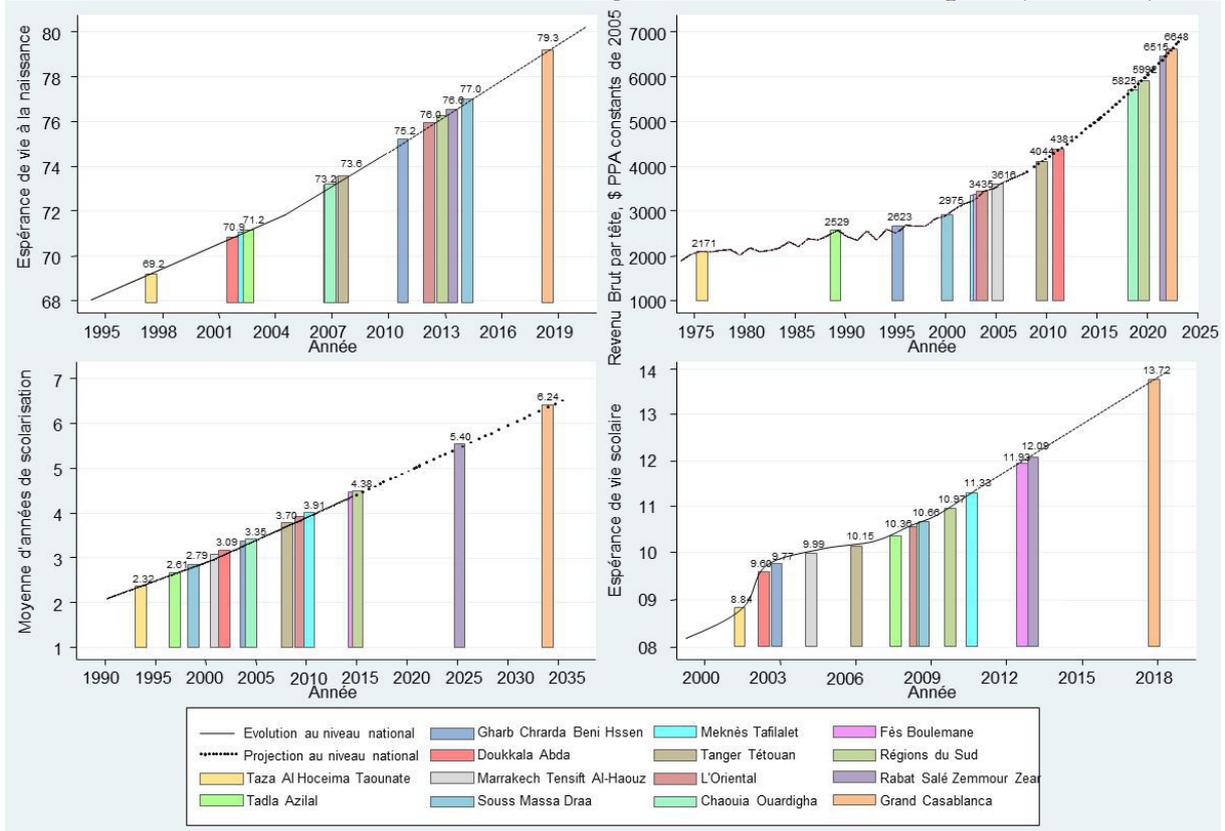
Annexe 5.3 : Fonction semi-paramétrique selon le niveau de développement et la zone géographique, indice Atkinson de l'éducation d'ordre 1



Annexe 6: Niveau de l'IDH non monétaire et des indices de santé et d'éducation par genre et milieu, par région (2010)



Annexe 7 : Retard en termes de composantes de l'IDH entre régions (en années)





Bibliographie

Bibliographie

- Adroque, C. (2013) "Equality of Educational Opportunities at Public Primary Schools in Argentina", *Education Policy Analysis Archives*, 21(89) : 1–25.
- Agrawal, T. (2014) "Educational Inequality in Rural and Urban India", *International Journal of Educational Development*, 34 : 11–19.
- Ahsan, H., M.E. Haque (2017) "Threshold Effects of Human Capital: Schooling and Economic Growth", *Economics Letters*, 156 : 48-52.
- Ahuja, V., D. Filmer (1996) "Educational Attainment in Developing Countries: New Estimates and Projections Disaggregated by Gender", *Journal of Educational Planning and Administration*, 10(3) : 229–254.
- Alesina, A., D. Rodrik (1994) "Distributive Politics and Economic Growth", *Quarterly Journal of Economics*, 109(2) : 465–490.
- Altinok, N. (2007) "Capital humain et croissance : l'apport des enquêtes internationales sur les acquis des élèves", *Economie Publique*, 18-19 2006/1-2 : 177-209.
- Anselin, L. (1996) "The Moran Scatterplot as an ESDA Tool To Assess Local Instability in Spatial Association", In: M. Fischer. H. Scholten and D. Unwin (eds.) *Spacial Analytical Perspectives on GIS in Environmental and Socio-Economic Scienses*. London: Taylor & Francis
- Anselin, L. (2010) "Thirty years of spatial econometrics", *Regional Science*, 89(1) : 3–25.
- Anselin, L., R.J.G.M. Florax, (eds.), *New Directions in Spatial Econometrics* (New York: Springer-Verlag, 1995)1, 75-95.
- Araujo, C., F. Ferreira, N. Schady (2004) "Is the World Becoming More Unequal? Changes in the World Distribution of Schooling", World Bank. Washington, DC.
- Ascher, W., N. Mirovitskaya (2016) "Social Services and Societal Fragility", Chapter 7 In: Ascher, W., N. Mirovitskaya, (eds.), *Development Strategies and Inter-Group Violence: Insights on Conflict-Sensitive Development*, 113-135.
- Ayesha, K., M. Kashif (2015) "The Impact of Educational and Gender Inequality on Income Inequality in South Asia", MPRA Paper 66661, University of Central Punjab.
- Baliamoune-Lutz, M., M. McGillivray (2009) "Does Gender Inequality Reduce Growth in Sub-Saharan African and Arab Countries?", *African Development Review*, 21(2) : 224–242.

- Baliamoune-Lutz, M., M. McGillivray (2015) "The Impact of Gender Inequality in Education on Income in Africa and the Middle East", *Economic Modelling*, 47 : 1–11.
- Banque asiatique de développement (2014) "Framework of Inclusive Growth Indicators 2014, 4th edition", Key Indicators for Asia and the Pacific Special Supplement
- Banque mondiale (2006): World Development Report 2006: Equity and Development. New York: Oxford University Press.
- Barro, R.J., J.W. Lee (2001) "International Data on Educational Attainment: Updates and Implications", *Oxford Economic Papers*, 53(3) : 541–563.
- Barro, R.J., J.W. Lee (2013) "A New Data Set of Educational Attainment in the World, 1950- 2010", *Journal of Development Economics*, 104(3) : 184–198.
- Barro, R.J., J.W. Lee. (1993) "International Comparisons of Educational Attainment", *Journal of Monetary Economics*, 32(3) : 363–394.
- Barro, R.J., J.W. Lee. (1996) "International Measures of Schooling Years and Schooling Quality", *American Economic Review*, 86(2) : 218–223.
- Bartusevičius, H. (2014) "The Inequality-Conflict Nexus Re-Examined: Income, Education and Popular Rebellions", *Journal of Peace Research*, 51(1) : 35–50.
- Becker, G.S. (1964) "Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education", New York: National Bureau of Economic Research.
- Benaabdelaali, W., S. Hanchane, A. Kamal (2012) Chapter 13 Educational Inequality in the World, 1950–2010: Estimates from a New Dataset, in John A. Bishop, Rafael Salas (ed.) *Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber* (Research on Economic Inequality, Volume 20), Emerald Group Publishing Limited, pp.337-366.
- Benaabdelaali, W., S. Hanchane, A. Kamal (2012) Human Capital Inequality and Economic Development: New Evidence from Nonparametric, and Semiparametric Panel Data Models, 1955-2010. In: The 4th Development Conference of the GREThA/GRES, Bordeaux.
- Benaabdelaali, W., Z. Bennaghmouch, S. Cheddadi, A. Kamal (2013) "Les disparités régionales de développement humain au Maroc", *Région et Développement*, 37 : 65–93.
- Benabou, R. (2000) "Unequal Societies: Income Distribution and the Social Contract", *American Economic Review*, 90(1) : 96–129.
- Benhabib J., M. Spiegel (1994) "The role of human capital in economic development. Evidence from aggregate cross-country data", *Journal of monetary Economics*, 34(2) :143-173.

Benos, N., S. Zotou (2014). "Education and economic growth: A meta-regression analysis", *World Development*, 64: 669-689.

Bernauer, E. (2016) "Ethnic, Religious and Class-Based Civil Conflicts", Chapter 2 In: Bernauer, E. (eds.), *Identities in Civil Conflict: How Ethnicity, Religion and Ideology Jointly Affect Rebellion* 25-40.

Berthélemy, J.C. (2002) "Convergence Clubs and Underdevelopment Traps", in *Development is Back*, édité par J. Braga de Macedo, C. Foy et Ch. Oman. Paris, OCDE Development Centre Studies, 61–76.

Berthélemy, J.C. (2006) "To What Extent are African Education Policies Pro-poor?", *Journal of African Economies*, 15(3) : 434–469.

Berthélemy, J.-C., F. Arestoff (2003) "Les stratégies d'éducation et le développement en Afrique", In : P. Bauchet et P. Germain, *L'éducation, fondement du développement durable en Afrique*. Paris, 95–132. (Cahier de sciences morales et politique, 1).

Betts J.R., K.S. Rueben, A. Danenberg (2001) "Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California", San Francisco: Public Policy Institute of California.

Birdsall, N., J.L. Londoño (1997) "Asset inequality matters: an assessment of the World Bank's approach to poverty reduction", *American Economic Review*, 87(2) : 32–37.

Borooah, V.K, C. Knox (2015) "Inequality and Segregation Differences in Educational Outcomes Between Grammar and Secondary Schools", In Borooah, V.K, C. Knox (eds.), *The Economics of Schooling in a Divided Society : the Case for Shared Education*. Chapter 5, 84–115.

Bosancianu, C. M. (2017). A Growing Rift in Values? Income and Educational Inequality and Their Impact on Mass Attitude Polarization*. *Social Science Quarterly*.

Bowman, K.J. (2007) "Knowledge Stocks by Distance to Frontier: Linking Low Education Inequality to High Growth in Developing Countries", *Journal of Asian Economics*, 18(4): 613–635.

Burt M.E. and N. Park (2009). "Education Inequality in the Republic of Korea: Measurement and Causes", in D. B. Holsinger, and W. J. Jacob (eds.), *Inequality in Education: Comparative and International Perspectives*.

Castelló, A. (2008) "On the distribution of education and democracy", *Journal of Development Economics*, 87 : 179–190.

Castelló, A. (2010a) "Channels Through Which Human Capital Inequality Influences Economic Growth", *Journal of Human Capital*, 4(4) : 394–450.

Castelló, A. (2010b) "Inequality and Growth in Advanced Economies: An Empirical Investigation", *Journal of Economic Inequality*, 8(3) : 293–321.

Castelló, A., A. Mukhopadhyay (2013) "Mass Education or a Minority Well Educated Elite in the Process of Growth: The Case of India", *Journal of Development Economics*, 105 : 303–320.

Castelló, A., R. Doménech (2002) "Human Capital Inequality and Economic Growth: Some New Evidence", *The Economic Journal*, 112(478) : 187–200.

Catin, M., A. Dieng, I. Sy (2015) "Disparités régionales et déterminants des performances et des inégalités scolaires dans les pays d'Afrique australe et de l'Est", XXXI^è Journées du développement ATM, Rouen, 3-5 juin 2015.

Checchi, D. (2004) "Does educational achievement help to explain income inequality?", In A. Cornia (eds.), *Inequality, Growth and Poverty in an Era of Liberalization and Globalization*. (Oxford University Press), Chapter 4.

Checchi, D., C. García-Peñalosa (2004) "Risk and the distribution of human capital", *Economics Letters* 82 : 53–61.

Chow, G.C, Y. Shen (2014) "Demand fo Education in China", In G. G Liu, S. Zhang, & Z. Shang (eds.), *Investment and human capital for economic development in china*, pp. 87-104. Singapore: World Scientific.

Cingano, F. (2014) "Trends in Income Inequality and Its Impact on Economic Growth", OECD/SEM Working Paper n°163, OECD Publishing.

Clarke, G.R.G. (1995) "More Evidence On Income Distribution and Growth", *Journal of Development Economics*, 47(2): 403-427.

Cleveland, W.S., (1979) "Robust Locally Weighted Regression and Smoothing Scatterplots", *Journal of the American Statistical Association*, 74 : 829–836.

Cleveland, W.S., S.J. Devlin (1988) "Locally Weighted Regression : An Approach to Regression Analysis by Local Fitting", *Journal of the American Statistical Association*, 83 : 596–610.

Cohen, D., M. Soto (2007) "Growth and Human Capital: Good Data, Good Results", *Journal of Economic Growth*, 12(1) : 51–76.

Collins, J.M. (2009) "Reconstructing Access in the Cambodian Education System", in Holsinger, D.B. and Jacob, W.J. (eds.), *Inequality in Education: Comparative and International Perspectives*. The Hague: Springer Netherlands.

Corcoran, S.P., W.N. Evans (2015) "Equity, Adequacy and the Evolution State Role in Education Finance", In Helen F. Ladd, Edward B. Fiske (eds.), *Handbook of Research in Education Finance and Policy*. Chapter 19, 332–356.

Cruces, G., C. G., Domenech, L. Gasparini (2014). "Inequality in education: evidence for Latin America", In G.A. Cornia (eds.) *Falling inequality in Latin America Policy changes and lessons*, 318-339.

- Cuaresma C.J., S. KC, P. Sauer (2013) "Age-Specific Education Inequality, Education Mobility and Income Growth"
- de la Croix, D., M. Doepke (2003) "Inequality and Growth: Why Differential Fertility Matters", *American Economic Review*, 93 : 1091–1113.
- De La Fuente, A., R. Doménech (2000) "Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?", CEPR Discussion Paper 2466. London: Centre for Economic Policy Research.
- De La Fuente, A., R. Doménech (2006) "Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?", *Journal of the European Economic Association*, 4(1) : 1–36.
- De La Fuente, A., R. Doménech (2015) "Educational Attainment in the OECD, 1960–2010. Updated Series and a Comparison With Other Sources", *Economics of Education Review* 48 : 56–74.
- Deininger, K., L. Squire (1996) "A New Data Set Measuring Income Inequality", *The World Bank Economic Review*, 10(3): 565-591.
- Delgado, M., D. J. Henderson C. Parmeter (2014). "Does Education Matter for Economic Growth?", *Oxford Bulletin of Economics and Statistics*, 76 (3) : 334-359.
- Dessus, S. (2000) "Capital humain et croissance : le rôle retrouvé du système éducatif", *Economie publique* 6(2): 95-114, dans Efficacité des systèmes éducatifs et de formation, coordonné par S. Hanchane.
- Dessus, S. (2001) "Human Capital and Growth: the Recovered Role of Educational Systems", World Bank Policy Research Working Paper, 2632 Washington, DC.
- Devarajan, S., V. Swaroop, H. Zou (1996) "The Composition of Public Expenditure and Economic Growth", *Journal of Monetary Economics*, 37 : 313–344.
- Digdowiseiso, K. (2009) "Education Inequality, Economic Growth, and Income Inequality: Evidence from Indonesia, 1996-2005", MPRA Paper No. 17792, University Library of Munich, Germany.
- Digdowiseiso, K. (2010) "Measuring Gini coefficient of Education: the Indonesian cases", MPRA Paper No. 19865, University Library of Munich, Germany.
- Duarte, A., M. Simões (2010) "Regional Growth in Portugal: Assessing the Contribution of Earnings and Education Inequality", Faculdade de Economia da Universidade de Coimbra,
- Easterly, W., N. Loayza, P. Montiel (1997) "Has Latin America's Post Reform Growth Been Disappointing?", *Journal of International Economics*, 43 : 287–311.

- Eberstadt, N., (2017) "Longevity, Education, and the Huge New Worldwide Increases in Equality.", In: J.P. Delsol, N. Lecaussin and E. Martin (eds.), *Anti-Piketty: Capital for the 21st Century* Section 1.2
- Felice E., Vasta M., 2012, "Passive Modernization? The New Human Development Index and Its Components in Italy's Regions (1871-2007)", UHE Working Paper, 2012.
- Felouzis G., S. Charmillot, B. Fouquet-Chauprade (2011) "Les inégalités scolaires en Suisse et leurs déclinaisons cantonales : l'apport de l'enquête Pisa 2003", *Swiss Journal of Sociology*, 37(1) : 33–55.
- Fidalgo, J.G., M.C.N. Simões, A. Duarte (2010) "Mind the Gap: Education Inequality at the Regional Level in Portugal, 1986-2005", *Notas Económicas* 32 : 22–43.
- Földvári, P., B. Van Leeuwen. (2010) "Should Less Inequality in Education Lead to a More Equal Income Distribution?", *Education Economics*, 1469–5782.
- Földvári, P., B. Van Leeuwen. (2011). "Should Less Inequality in Education Lead to a More Equal Income Distribution?", *Education Economics*, 19(5): 537–554.
- Földvári, P., B. Van Leeuwen (2014) "Educational and income inequality in Europe, ca. 1870–2000", *Cliometrica*, 8(3) : 271–300.
- Forbes, K. J. (2000) "A reassessment of the relationship between inequality and growth", *American Economic Review* 90(4): 869–887.
- Galea, S., J. Ahern, M. Tracy, S. Rudenstine, D. Vlahov (2007) "Education Inequality and Use of Cigarettes, Alcohol, and Marijuana", *Drug and Alcohol Dependence*, 90(1) : 4–15.
- Galor, O., J. Zeira (1993) "Income Distribution and Macroeconomics", *Review of Economic Studies*, 60 : 35–52.
- Galor, O., O. Moav (2004) "From Physical to Human Capital Accumulation: Inequality and the Process of Development", *Review of Economic Studies*, 71 : 1001–1026.
- Garces-Voisinat, J. P. (2016) "Equality of Opportunity in Education: A Case Study of Chile and Norway", *Journal of Economics and Political Economy*, 3(1): 142-150.
- Gille, V. (2015) "Distribution of Human Capital and Income: An Empirical Study on Indian States", *Journal of Macroeconomics*, 43 : 239–256.
- Giorgi, G.M., C. Gigliarano (2017) "The Gini Concentration Index: A Review of the Inference Literature", *Journal of Economic Surveys*, 31(4): 1130–1148.
- Green, A. (2011) "Lifelong Learning, Equality and Social Cohesion", *European Journal of Education*, 46(2): 228–243.

Green, A., J. Preston, J.G. Janmaat (2006) "Education, Equality and Social Cohesion: a comparative analysis", Basingstoke: Palgrave Macmillan.

Gregorio, J.D., J.W. Lee (2002) "Education and Income Inequality: New Evidence From Cross-Country Data", *Review of Income and Wealth*, 48(3) : 395–416.

Güngör, N.D. (2010) "Education, Human Capital Inequality and Economic Growth: Evidence from Turkey", *Regional and Sectoral Economic Studies*, 10(2): 53–71.

Gurgand, M. (2000) "Capital humain et croissance : la littérature empirique à un tournant ?", *Economie publique* 6(2): 71-93, dans Efficacité des systèmes éducatifs et de formation, coordonné par S. Hanchane.

Gustave, D.N, N.P. Joubert (2012) "Analysis of the Evolution of Education Inequality in Cameroon", 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong

Halter D., M. Oechslin, J. Zweimüller, (2014) "Inequality and growth: the neglected time dimension", *Journal of Economic Growth*, 19(1): 81-104.

Hassan, R., M. Shahzad (2007) "Education Inequality and Economic Growth: Framework for the Evaluation of Pakistan's Education Policy", *International Journal of Human Development*, 3(1) 37–60.

Hazell, E., K.F. Gee, A. Sharpe (2012) "The Human Development Index in Canada: Estimates For The Canadian Provinces And Territories, 2000-2011", CSLS Research Report, 02.

Henderson, D.J., R.J. Carroll Q. Li (2008) "Nonparametric Estimation and Testing of Fixed Effects Panel Data Models", *Journal of Econometrics*, 144 : 257–275.

Holsinger, D.B (2009) "The Distribution of Education in Vietnam: Why Does Equality Matter?", In Y. Hirotsu, Y. Kitamura (eds.), *The Political Economy of Educational Reforms and Capacity Development in Southeast Asia*, pp. 191-216.

Holton, C.J. (2012) "What are the effects of fossil-fuel subsidies on growth, the environment, and inequality?", the School of Economics, University of Nottingham.

Hori, T. (2011) "Educational Gender Inequality and Inverted U-Shaped Fertility Dynamics", *The Japanese Economic Review* 62(1) : 126–150.

Iatarola, P., L. Stiefel (2003) "Intradistrict Equity of Public Education Resources and Performance", *Economics of Education Review* 22(1) : 69–78.

Ibourk, A., J. Amaghous (2013) "Inequality in Education and Economic Growth: Empirical Investigation and Foundations - Evidence from MENA Region", *International Journal of Economics and Finance*, 5(2) : 111–124.

- Ilon, L. (2011) "Can education equality trickle-down to economic growth? The case of Korea", *Asia Pacific Education Review*, 12(4) 653–663.
- Islam, N., (1995) "Growth Empirics : A Panel Data Approach", *Quarterly Journal of Economics*, , 110 (4): 1127-1170.
- Jamal H., A.J. Khan (2005) "Impact of Ownership and Concentration of Land on Schooling", *Lahore Journal of Economics*, 10(2) : 1–14.
- Johnes, G. (2006) "Education and Economic Growth", Working Paper 2006/019 Lancaster University Management School
- Jordá, V. (2015) "The Global Distribution of Education: 1970-2010", 6th Meeting of the Society for the Study of Economic Inequality ECINEQ Université du Luxembourg, Luxembourg, July 13-15, 2015
- Jordá, V., J. M. Alonso (2017) "New Estimates on Educational Attainment Using a Continuous Approach (1970–2010)", *World Development*, 90: 281-293.
- Kaldor N. (1956) "Alternative Theories of Distribution", *Review of Economic Studies*, 23(2): 94-100.
- Klasen, S., F. Lamanna (2009) "The Impact of Gender Inequality in Education and Employment on Economic Growth: New Evidence for a Panel of Countries", *Feminist Economics*, 15(3) : 91–132.
- Kochav, Y., A. Spivak, M. Strawczynski "Human Capital, Educational Levels and Growth", presented at XXIV Meeting of the Economics of Education Association in 2015 June Madrid <http://2015.economicsofeducation.com/user/pdfsiones/085.pdf>
- Krueger, A.B., M. Lindahl (2001) "Education for Growth: Why and For Whom?", *Journal of Economic Literature*, 39(4) : 1101–1136.
- Kuznets, S. (1955) "Economic Growth and Income Inequality", *American Economic Review*, 65 : 1-28.
- Kyriacou, G.A. (1991) "Level and Growth Effects of Human Capital: A Cross–Country Study of the Convergence Hypothesis", Economic Research Reports: 19–26, C.V. Starr Center for Applied Economics, New York University.
- Lambert, P.J. (2001) "The Distribution and Redistribution of Income", (3rd ed.). Manchester: Manchester University Press.
- Lau, L.J., D. T. Jamison, F. F. Louat (1991) "Education and Productivity in Developing Countries: An Aggregate Production Function Approach", World Bank PRE Working Paper Series 612, Washington D.C.
- Lê, F., J. Ahern, S. Galea (2010) "Neighborhood education inequality and drinking behavior", *Drug and Alcohol Dependence* 112(1) : 18–26.
- Li H., L., Squire H. Zou (1998) "Explaining International and Intertemporal Variations in Income Inequality", *Economic Journal*, 108: 26-43.

- Li, Q., A. Ullah (1998) "Estimating Partially Linear Panel Data Models with One-way Error Components", *Econometric Reviews*, 17(2) : 145–166.
- Li, Q., T. Stengos (1996) "Semiparametric Estimation of Partially Linear Regression Models", *Journal of Econometrics*, 71 : 389–397.
- Lim, A.S.K., K.K. Tang (2008) "Human Capital Inequality And The Kuznets Curve", *The Developing Economies*, 46(1) : 26–51.
- Lin, C.H.A, C.H. Yang (2009) "An Analysis of Educational Inequality in Taiwan After the Higher Education Expansion", *Social Indicators Research*, 90(2) : 295–305.
- Lin, X., R.J. Carroll (2006) "Semiparametric Estimation in General Repeated Measures Problems", *Journal of the Royal Statistical Society*, 68 : 68–88.
- Lipták, K. (2010) "Representation of Modified Human Development Index in Romania's and Hungary's Nuts2 Regions", *Managerial Challenges of the Contemporary Society*, 1 : 92–96.
- Londoño, J.L. (1990) "Kuznetsian Tales with Attention to Human Capital", Seminar paper at the Third Inter-American Seminar in Economics, Rio de Janeiro, Brazil.
- Longley, P. A., C. Tobon (2004) "Spatial Dependence and Heterogeneity in Patterns of Hardship: An Intra-urban Analysis", *Annals of the Association of American Geographers*, 94 : 503–519.
- López, R., V. Thomas, and Y. Wang (1998). "Addressing the Education Puzzle: the Distribution of Education and Economic Reforms.", World Bank Working Paper Series No. 2031, World Bank, Washington, DC.
- Lorel, B. (2008). "Assessing Brazilian Educational Inequalities.", *Revista Brasileira de Economia*, 62(1): 31–56.
- Lutz, W., A. Goujon, K. C. Samir, W. Sanderson (2007) "Reconstruction of populations by age, sex and level of educational attainment for 120 countries for 1970–2000", *Vienna Yearbook of Population Research* 2007 : 193–235.
- Lv, X., R. Li, Z. Fang, (2017) "Efficient Semiparametric Estimation for Gini Inequality Treatment Effects", *Economics Letters*, 154 : 96-100.
- Mankiw, N.G., D., Romer, D. Weil, (1992) "A contribution to the empirics of economic growth", *Quarterly Journal of Econometrics*, CVII, 407-437.
- Marsh, R. M., (2016) "What Have We Learned from Cross-National Research on the Causes of Income Inequality?", *Comparative Sociology* 15: 7-36.
- Martins, S. (2016). "Performance, Organisation and Inequality in Education Systems: An International Comparison of the Dynamics of Schooling", *International E-Journal of Advances in Education*, 2(6) : 374-382.

- Masakure, O. (2013) "Education Inequality and Economic Development in Eastern and Southern Africa", Chapitre 7, E. Shizha and A. Abdi (eds) *Indigenous Discourses and Knowledge in Africa*. Taylor & Francis/Routledge.
- McKenzie, D., H. Rapoport (2007) "Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico", *Journal of Development Economics* 84(1) : 1–24.
- Merle, P. (2002) "Démocratisation ou accroissement des inégalités scolaires? L'exemple de l'évolution de la durée des études en France (1988-1998)", *Population*, 57(4) : 633–659.
- Mesa, E.P. (2007) "Measuring education inequality in the Philippines", *Philippine Review of Economics*, 44(2) : 33–70.
- Meschi, E. F. Scervini (2014) "A New Dataset on Educational Inequality", *Empirical Economics*, 47(2) : 695–716.
- Monchar, P.H. (1981) "Regional Educational Inequality and Political Instability", *Comparative Education Review* 25(1) : 1–12.
- Morales, M., C.P. Terán (2010) "Educational Inequality in Argentina: The Best and Worst Performers", Documentos de Trabajo n° 5, Universidad Nacional de Salta.
- Morrisson, C., F. Murtin (2009) "The Century of Education", Centre for Economic Performance, CEP Discussion Paper No. 934.
- Morrisson, C., F.Murtin (2010) "The Kuznets Curve of Education: A Global Perspective on Education Inequalities", Centre for the Economics of Education, London School of Economics, CEE DP 116.
- Morrisson, C., F.Murtin (2013) "The Kuznets curve of human capital inequality: 1870–2010", *Journal of Economic Inequality*, 11(3) : 283–301.
- Müller, W. (2014) "Educational Inequality and Social Justice: Challenges for Career Guidance", *International Journal for Educational and Vocational Guidance*, 14 (1) : 21–33.
- Nadaraya, E. (1964) "On Estimating Regression. *Theory of Probability and its Applications*", 9 : 141–142.
- Nehru, V., E. Swanson, A. Dubey (1995) "A New Database on Human Capital Stock: Sources, Methodology and Results", *Journal of Development Economics* 46(2) : 379–401.
- Nations Unies (2013) "Inequality Matters", Report on the World Social Situation <http://www.un.org/esa/socdev/documents/reports/InequalityMatters.pdf>
- Ostry J., A. Berg, C. Tsangarides (2014) "Redistribution, Inequality, and Growth", IMF Staff discussion note, February

- Oyvatt, C. (2016). "Agrarian Structures, Urbanization, and Inequality" *World Development*, 83: 207-230.
- Park, J. (2006). "Dispersion of Human Capital and Economic Growth.", *Journal of Macroeconomics*, 28 : 520–539.
- Patrawart, K. (2010) "Can Equality in Education Be a New Anti-Corruption Tool? Cross-Country Evidence (1990-2005)", *NACC Journal*, 3(2) : 127–157.
- Persson, T., G. Tabellini (1994) "Is Inequality Harmful for Growth?", *American Economic Review*, 84(3) : 600–621.
- Peterson, L. (2014). "The measurement of non-economic inequality in well-being indices.", *Social indicators research*, 119(2) : 581-598.
- PNUD (2010) "Rapport sur le développement humain 2010. La vraie richesse des nations : Les chemins du développement humain", Programme des Nations Unies pour le développement
- Potancokova, M., S. KC, A. Goujon, R. Bauer (2014) "New Global Estimates of Mean Years of Schooling for 171 Countries", European Population Conference 2014
- Potter, M. (2014). "Picking apart Piketty". *Policy: A Journal of Public Policy and Ideas*, 30(4) : 36-47.
- Pritchett, L. 2001. "Where Has All the Education Gone ?", *World Bank Economic Review*, 15: 367-391.
- Psacharopoulos, G. (1977) "Unequal Access to Education and Income Distribution: An International Comparison", *De Economist* 125 : 383–392.
- Psacharopoulos, G., A. M. Arriagada (1986) "The Educational Composition of the Labour Force: An international Comparison", *International Labour Review*, 125(5) : 561–574.
- Qian, X., R. Smyth (2008) "Measuring Regional Inequality of Education in China: Widening Coast-Inland Gap or Widening Rural-Urban Gap?", *Journal of International Development* 20(1) : 132–144.
- Ram, R. (1982) "Public Subsidization of Schooling and Inequality of Educational Access: A New World-Cross Section Study", *Comparative Education Review*, 26 : 36–47.
- Ram, R. (1990) "Educational Expansion and Schooling Inequality: International Evidence and some Implications", *Review of Economics and Statistics*, 72(2) : 266–274.
- Rao, R., R.b., Jani (2008) "School Quality, Educational Inequality and Economic Growth", *International Education Studies*, 1(2) : 135–141.
- Rao, R., R.b. Jani (2011) "Teacher Allocation and Equity in Malaysian Schools", *International Journal of Institutions and Economies*, 3(1) : 103–112.

RNDH Ethiopie (2015) "Accelerating Inclusive Growth for Sustainable Human Development in Ethiopia", National Human Development Report, Programme des Nations Unies pour le développement.

RNDH Moldova (2014) "Moldova National Human Development Report", Programme des Nations Unies pour le développement.

RNDH Nepal (2014) "Nepal Human Development Report 2014 Beyond Geography Unlocking Human Potential", Programme des Nations Unies pour le développement.

RNDH Tanzania (2014) "Economic Transformation for Human Development", Tanzania National Human Development Report, Programme des Nations Unies pour le développement

Rodríguez-Poo, J.M., A. Soberon (2017) "Nonparametric and Semiparametric Panel Data Models: Recent Developments", *Journal of Economic Surveys*, 31(4): 923–960.

Rodríguez-Pose, A., V. Tselios (2010) "Inequalities in Income and Education and Regional Economic Growth in Western Europe", *The Annals of Regional Science*, 44(2) : 349–375.

Rosenblatt, M. (1956) "Remarks on Some Nonparametric Estimates of a Density Function", *Annals of Mathematical Statistics* 27(3) : 832–837.

Sabermahani A., M. Barouni, H. Seyedin, A. Aryankhesal (2013) "Provincial Human Development Index, a Guide for Efficiency Level Analysis: The Case of Iran", *Iranian Journal of Public Health*, 42(2) : 149–157.

Sahn D.E., S.D. Younger (2007) "Inequality and Poverty in Africa in an Era of Globalization Looking Beyond Income to Health and Education", Research Paper No. 2007/74, UNU World Institute for Development Economics Research.

Sahn, D.E., D.C. Stifel (2003) "Urban-Rural Inequality in Living Standards in Africa", *Journal of African Economies*, 12(4) : 564–597.

Sauer, P (2016) "The Role of Age and Gender in Education Expansion". INEQ Working Paper Series, 4. WU Vienna University of Economics and Business, Vienna.

Sauer, P., M. Zagler (2014) "(In)equality of Education and Economic Growth", *Review of Income and Wealth*, 60(2) : 353–379.

Schaffar, A. (2014) "Advances in Spatial Econometrics, Introduction to the Special Issue", *Région et Développement*, 40

Schrott L., M. Gachter, E. Theurl (2012) "Regional Development in Advanced Countries: A Within-Country Application of the Human Development Index for Austria", Working Papers in Economics and Statistics, 23.

Senadza, B. (2012) "Education Inequality in Ghana: Gender and Spatial Dimensions", *Journal of Economic Studies*, 39(6) 724–739.

- Sheret, M. (1984) "Note on Methodology: The Coefficient of Variation", *Comparative Education Review*, 28(3) : 467–476.
- Sherman J.D., J.M. Poirier (2007) "Educational equity and public policy: Comparing results from 16 countries", UNESCO Institution for Statistics.
- Silva, R., A.F. Lopes (2012) "A Regional Human Development Index for Portugal", CEFAGE-UE Working Paper, 05.
- Tan, S.C., C.M. Ho, V. Pang (2015) "Education inequality: become better or worse?", MPRA Paper No. 61627, University Library of Munich, Germany.
- Tansel, A., A.D. Güngör (2000) "Provincial Inequalities in School Enrollments in Turkey", Economic Research Forum Working Paper 2003.
- Theil, H. (1967) "Economics and Information Theory", North-Holland Publishing Company, Amsterdam.
- Thomas, V., Y. Wang, X. Fan (2001) "Measuring Education Inequality: Gini Coefficients of Education.", Policy Research Working Paper, N°. 2525, World Bank Institute
- Thomas, V., Y. Wang, X. Fan (2003) "Measuring Education Inequality: Gini Coefficients of Education for 140 Countries (1960–2000)", *Journal of Education Planning and Administration*, 17(1) : 5–33.
- Tomul, E. (2009) "Measuring Regional Inequality of Education in Turkey: An Evaluation by Gini Index", *Procedia Social and Behavioral Sciences*, 1 : 949–952.
- Trabelsi, S. (2013) "Regional Inequality of Education in Tunisia: An Evaluation by the Gini Index", *Region et Developpement*, 37 : 95–117.
- Umar, H.M., R. Ismail, R. Abdul-Hakim (2014a) "Regional Inequality of Educational Attainment in Nigeria", *British Journal of Economics Management & Trade*, 4(3) : 420–430.
- Umar, H.M., R. Ismail, L.H. Eam (2014b) "A Spatial Econometrics Analysis of Educational Distribution and Regional Income Disparities in Nigeria", *Handbook on the Emerging Trends in Scientific Research*, 1 : 722–731.
- Wang, N. (2003) "Marginal Nonparametric Kernel Regression Accounting for Within-Subject Correlation", *Biometrika*, 90 : 43–52.
- Watson, G. (1964) "Smooth Regression Analysis", *The Indian Journal of Statistics*, 26 : 359–372.
- Yang, J., X. Huang, X. Li (2014) "An analysis of education inequality in China", *International Journal of Educational Development* 37 : 2–10.
- Yang, J., X.S. Li (2007) "Education Inequality, Human Capital and Economic Growth: An Empirical Study on China", *The Journal of Quantitative & Technical Economics*, 2 : 37–45.

Yu, N., B. Yu M. de Jong S. Storm (2015) "Does Inequality in Educational Attainment Matter for China's Economic Growth?", *International Journal of Educational Development*, 41 : 164–173.

Zamora, C.M.B., R.A. Dorado (2015) "Rural-Urban Education Inequality in the Philippines Using Decomposition Analysis", *Journal of Economics, Management & Agricultural Development*, 1(1) : 63–71.

Zhang J., T. Li (2002) "International Inequality and Convergence in Educational Attainment, 1960–1990", *Review of Development Economics*, 6(3) : 383–392.

Zhang, C., K. Jin (2010) "Effect of Equity in Education on the Quality of Economic Growth: Evidence from China.", *International Journal of Human Sciences*, 7(1) : 47–69.

Zhou, X., K.W. Li (2011) "Inequality and Development: Evidence from Semiparametric Estimation With Panel Data", *Economics Letters*, 113 : 203–207.



Tables

Index des tableaux

Tableau 1: Formules des indicateurs d'inégalité d'éducation les plus utilisés	27
Tableau 2: Quelques travaux sur les inégalités éducatives selon divers variables d'éducation, entités d'observation, indices d'inégalité et populations d'intérêt	29
Tableau 3: Quelques travaux sur la décomposition des inégalités éducatives selon différents critères de décomposition.....	35
Tableau 4: Chronologie des bases mondiales sur les inégalités d'éducation..	37
Tableau 5: Chronologie des bases de données sur le niveau d'éducation maximum atteint.....	37
Tableau 6: Comparaison entre les bases de données de Cohen et Soto (2007).....	38
Tableau 7: Répartition de la population par tranche d'âge selon le niveau d'éducation maximum atteint pour le cas de la Bulgarie en 1990	44
Tableau 8: Moyenne d'année de scolarisation et nombre d'années cumulées de scolarisation par tranche d'âge pour le cas de la Bulgarie en 1990	44
Tableau 9: Indice de Gini de l'éducation par niveau de développement et par groupes d'âge, 1950-2010	63
Tableau 10: Exemple illustratif du problème d'agrégation	76
Tableau 11: Estimation paramétrique à effets fixes du modèle linéaire, quadratique et cubique du panel de pays en développement... 	94
Tableau 12: La partie paramétrique de la modélisation semi-paramétrique selon l'indice d'inégalité d'éducation utilisé.....	98
Tableau 13: Tableau synoptique des travaux concernant l'indice de Gini de l'éducation et de l'IDH au niveau infra-national	112
Tableau 14: Nombre de niveaux d'éducation et population d'intérêt retenus dans quelques travaux concernant l'IDH et le Gini d'éducation au niveau infra-national	114

Tableau 15: Nombre d'années de scolarisation correspondant à chaque niveau d'éducation selon l'ancienne et la nouvelle structure du système éducatif marocain	116
Tableau 16 : Indice de Gini et moyenne d'années de scolarisation par région.....	118
Tableau 17: Répartition de la population âgée de 15 ans et plus par niveau d'éducation atteint par région.....	120
Tableau 18 : Indice de Gini régional de l'éducation par genre	122
Tableau 19 : Indice de Gini régional de l'éducation par milieu rural/urbain	123
Tableau 20 : Indice de Gini régional selon le genre et le milieu de résidence.....	124
Tableau 21: Valeurs de l'IDH et de ses composantes, régions marocaines (2010)	126
Tableau 22: Rang de l'IDH et de ses composantes : classement national et international, régions marocaines (2010)	128
Tableau 23: Retard en développement humain par rapport à.....	131
Tableau 24: Indice de Gini de l'éducation et IDH des vingt premières et dernières communes, 2004.....	136
Tableau 25 : Statistique de Moran pour l'IDH et l'indice de Gini de l'éducation	138
Tableau 26 : Diagramme de Moran : récapitulation des associations	139
Tableau 27 : Résultats des estimations du modèle par les MCO et du modèle à autocorrélation spatiale des termes d'erreur	144
Tableau 28: Résultats des estimations du modèle par les MCO et du modèle à autocorrélation spatiale des termes d'erreur.....	146
Tableau 29: Estimation paramétrique linéaire à effets fixes du panel provincial	150
Tableau 30: Estimations paramétrique et semi-paramétrique à effets fixes du panel provincial	152

Index des graphiques

Graphique 1 : Inégalité d'éducation et développement économique, cas du Nicaragua et du Vietnam.....	15
Graphique 2 : Courbe de Lorenz de l'éducation et l'interprétation géométrique de l'indice de Gini de l'éducation	32
Graphique 3 : Contribution graphique de l'analphabétisme à la surface d'inégalité éducative	54
Graphique 4 : Courbe de Lorenz pour le cas de quatre niveaux d'éducation	56
Graphique 5 : Courbe de Lorenz pour le cas de "n" niveaux d'éducation.....	58
Graphique 6 : Domaine de variation de l'indice de Gini de l'éducation selon la moyenne d'années de scolarisation.....	62
Graphique 7 : Indice de Gini de l'éducation par niveau de développement et groupes d'âge, 1950-2010.....	64
Graphique 8 : La courbe de Lorenz de l'éducation de la population âgée de 15 ans et plus par niveau de développement, 1950-2010	65
Graphique 9 : La courbe de Lorenz de l'éducation par groupes d'âge et par niveau de développement, 1950-2010	66
Graphique 10 : indice de Gini de l'éducation par genre, niveau de développement et zone géographique, 1950-2010	67
Graphique 11 : lien entre l'indice de Gini de l'éducation et la moyenne d'année de scolarisation par zone géographique, 1950-2010.....	68
Graphique 12 : Lien entre l'écart type de scolarisation et la moyenne d'années de scolarisation par zone géographique, 1950-2010.....	69
Graphique 13 Le cas particulier de l'indice Atkinson d'ordre 1.....	84
Graphique 14 Fonction non-paramétrique estimée pour le panel des pays en développement selon l'indice d'inégalité d'éducation	96
Graphique 15 Fonction semi-paramétrique pour le panel des pays en développement selon l'indice d'inégalité d'éducation	97
Graphique 16 Illustration de la présence de clubs de convergence dans la relation éducation-croissance.....	99
Graphique 17 Fonction non-paramétrique selon le niveau de développement et a zone géographique, Indice de Gini de l'éducation	100
Graphique 18 Schéma synthétique des phases de la relation inégalités éducatives / développement économique	101

Graphique 19 Fonction semi-paramétrique selon le niveau de développement et la zone géographique, indice de Gini de l'éducation	103
Graphique 20. Décomposition de l'effet des inégalités éducatives sur le développement économique: inégalités intra-éduqués et inter éduqués-illettrés.....	104
Graphique 21 Dépendance fonctionnelle liant le taux d'analphabétisme aux inégalités entre individus éduqués et illettrés selon différents indices d'inégalité.....	105
Graphique 22 Décomposition de l'effet des inégalités éducatives sur le développement économique : inégalités intra-générationnelles et intergénérationnelles	106
Graphique 23 : Retard de développement entre régions (en années).....	130
Graphique 24 Rang des régions marocaines dans le classement mondial de l'indice de Gini de l'éducation et de l'IDH, 2010.....	131
Graphique 25: Diagramme de Moran de l'indice de développement humain et de l'indice de Gini de l'éducation	139
Graphique 26 : Illustration de la trajectoire du lien entre inégalités dans l'éducation et développement	145
Graphique 27 : Fonctions non-paramétriques estimées pour le panel des pays	151
Graphique 28 : Décomposition de l'effet des inégalités éducatives sur le développement des provinces: inégalités intra-éduqués et inter éduqués-illettrés ..	154
Graphique 29: Décomposition de l'effet des inégalités éducatives sur le développement des provinces : inégalités intra-générationnelles et intergénérationnelles	155

Index des figures

Figure 1 : Schéma de différenciation des entités d'observation selon un critère donné de décomposition (cas d'un critère de trois groupes).....	33
Figure 2 : Indice de Gini de l'éducation de la population âgée de 15 ans et plus à l'échelle mondiale entre 1950 et 2010.....	63
Figure 3 : Indice de Gini de l'éducation et de l'IDH dans les communes marocaines, 2004.....	137
Figure 4 : LISA pour l'indice de développement humain et l'indice de Gini.....	140

Abstract

This thesis seeks to deepen the nature and the shape of the relationships between educational inequalities and development. It goes along with the prolongation of the analyses undertaken about the relationships between both education & growth; and inequality & growth, by trying to shed additional light on these two. It aims to retrace, in a schematic way, the evolution of educational inequalities particularly in the developing countries; and also to characterize the nonlinearity of this link using nonparametric and semiparametric estimation models.

This thesis consists of three chapters that correspond to specific objectives, databases and methodologies. In the first chapter, we propose a new dataset on the inequalities of education. Most of the studies on the relationship between human capital and economic development have mainly apprehended the measurement of human capital through quantitative education indicators, using namely the average of years of schooling (human capital stock). Our database, which presents an alternative measure of human capital, tends to improve significantly the way in which inequalities in education are calculated. It employs all the abundance of disaggregated data, corrects inappropriate weightings and refines some reductive assumptions about the durations of schooling cycles and the levels of education. We have also generalized the formula proposed by Berthélemy (2006) on the Gini index of education. The possible variation range of this index is graphically identified according to the average years of schooling and the cumulative duration of the schooling cycles.

In Chapter II, we reveal the existence of a nonlinear relationship between inequalities in education and economic development using nonparametric and semiparametric models that do not require predefined functional forms. Several phases are therefore highlighted: the first three are identified only according to the level of development; then two other phases are recognized by combining thresholds of both development and education inequality; the sixth and final phase is defined by the educational inequality level alone. We show that the phases during which the reduction of educational inequality presents the most beneficial impact on economic development are the third and the fifth.

Beyond the general outline highlighted at the transnational level in Chapters I and II, we explore in Chapter III the nature of this relationship at the regional level in the case of Morocco, for which we have data at both the municipal and provincial levels. We also confirm the nonlinearity of the relationship. The third phase, identified in Chapter II, is divided to two sub-phases in the case of the Moroccan provinces which have a differentiated impact according to a threshold of development and inequality of education.

Résumé

Cette thèse cherche à approfondir la nature et la forme des relations entre les inégalités éducatives et le développement. Elle s'inscrit dans le prolongement des analyses engagées sur les liens éducation/croissance et inégalités/croissance, en essayant d'apporter un éclairage complémentaire sur ces deux relations. Elle vise à retracer de manière stylisée l'évolution des inégalités éducatives particulièrement dans les pays en développement et à caractériser la non-linéarité de la relation à partir de l'estimation de modèles non-paramétriques et semi-paramétriques.

Cette thèse est constituée de trois chapitres auxquels correspondent des objectifs, des bases de données et des méthodologies spécifiques. Dans un premier chapitre, nous proposons une nouvelle base mondiale sur les inégalités d'éducation. La majorité des travaux sur la relation entre capital humain et développement économique ont principalement appréhendé la mesure du capital humain à travers des mesures de l'éducation en utilisant notamment la moyenne d'années de scolarisation (stock du capital humain). Notre base de données, qui présente une mesure alternative du capital humain, tend à améliorer sensiblement le mode de calcul des inégalités de l'éducation. Elle exploite toute la richesse des données désagrégées, corrige les pondérations inappropriées et affine certaines hypothèses réductrices sur les durées des cycles d'enseignement et les niveaux d'éducation retenus. Nous avons aussi généralisé la formule proposée par Berthélemy (2006) sur l'indice de Gini de l'éducation. Le domaine de variation possible de cet indice est identifié graphiquement selon la moyenne d'années de scolarisation et les durées cumulées des cycles d'enseignement.

Nous mettons en évidence, dans le cadre du chapitre II, l'existence d'une relation non linéaire entre les inégalités dans l'éducation et le développement économique en utilisant des modèles non-paramétriques et semi-paramétriques qui n'exigent pas de formes fonctionnelles prédéfinies à l'avance. Plusieurs phases sont ainsi mises en évidence : les trois premières sont repérées seulement par rapport aux niveaux de développement ; deux autres sont identifiées à la fois par des seuils de développement et d'inégalité d'éducation ; une sixième et dernière phase est définie par rapport au seul niveau d'inégalité d'éducation. Nous montrons que c'est dans la troisième et cinquième phases que la réduction de l'inégalité d'éducation présente l'impact le plus bénéfique sur le développement économique.

Au-delà du schéma général mis en évidence sur le plan transnational dans les chapitres I et II, nous explorons dans le chapitre III la nature de cette relation au plan régional dans le cas du Maroc, pour lequel nous disposons de données aux niveaux communal et provincial. La non-linéarité de la relation est aussi confirmée. La troisième phase repérée au chapitre II est subdivisée, dans le cas des provinces marocaines, en deux sous phases qui présentent un impact différencié selon un seuil de développement et d'inégalité d'éducation.