



Munich Personal RePEc Archive

Contribution of unobservables to gender disparities in schooling and child labor in Mali

Keita, Moussa

24 July 2014

Online at <https://mpra.ub.uni-muenchen.de/57532/>
MPRA Paper No. 57532, posted 24 Jul 2014 21:57 UTC

Contribution des inobservables aux disparités de genre dans la scolarisation et le travail des enfants au Mali

Moussa Keita

Doctorant

Centre d'Etudes et de Recherches sur le Développement International

(Juillet 2014)

Résumé

Cette étude tente de prolonger les réflexions sur les disparités de genre dans l'allocation du temps des enfants en s'appuyant sur les données d'enquête ménages au Mali¹. Nous testons, en particulier, les disparités de genre dans la scolarisation et le travail des enfants en utilisant la décomposition Oaxaca-Blinder. L'objectif étant d'identifier les contributions des facteurs inobservables aux différences entre filles et garçons. Nos résultats confortent, en effet, l'hypothèse selon laquelle l'essentiel des différences entre les deux groupes provient des facteurs inobservables. Nous trouvons, par exemple, que 92% de la différence dans la scolarisation reste non expliquée à partir des caractéristiques observables. Nous mettons également en évidence une forte polarisation des travaux selon le genre. Les garçons sont significativement plus orientés vers les activités économiques (1 heure de plus que les filles par jour) alors que les filles sont significativement plus orientées vers les activités non économiques (une différence avec les garçons d'environ 2.5 heures par jour).

Codes JEL: C12 D13 I21 J22

Mots clés: Scolarisation, travail des enfants, disparités de genre

Contact Info

Email : Moussa.keita@udamail.fr ou keitam09@ymail.com

¹ Les données de cette étude proviennent d'une enquête ménages réalisée par le CERDI en collaboration avec l'Institut d'Economie Rurale (Bamako, Mali) sur financement du Programme Régional Energie pour la réduction de la Pauvreté du PNUD (PNUD/PREP) à Dakar (Sénégal). L'auteur tient à remercier l'ensemble de ces structures pour la mise à disposition des données. Bien entendu, les idées et opinions émises n'engagent en rien la responsabilité de ces organismes. Ces idées et opinions ainsi que les erreurs et insuffisances de ce manuscrit doivent être strictement considérées comme propres à leur auteur.

Introduction

Plusieurs études tendent à montrer que les comportements des parents en termes de scolarisation et de mise au travail des enfants diffèrent très significativement selon que l'enfant soit une fille ou un garçon. Il ressort de nombreuses études l'idée selon laquelle les parents ne font pas les mêmes choix pour leur enfant selon son genre. Les pères s'occupent davantage des garçons et les mères s'occupent davantage des filles (voir Ridao-Cano, 2000 ; Basu, 2001 ; Wahba, 2005 ; Emerson et Souza, 2007). Par exemple, Emerson et Souza (2007) trouvent sur les données du Brésil que le niveau d'éducation des pères influence la scolarisation et le travail des garçons significativement plus que ceux des filles, tandis que le niveau d'éducation des mères, lui, influence significativement le travail des filles plus que celui des garçons. Par ailleurs, d'autres études montrent une répartition des travaux selon le genre très marquée. Par exemple, Koissy-Kpein(2012) montrent que les filles sont plus significativement orientées vers les travaux « non-économiques » alors que les garçons sont significativement orientés vers les travaux « économiques ». En plus de cette répartition des travaux selon le genre, il est aussi largement reconnu que les garçons ont une probabilité plus forte d'être scolarisés que les filles. Tous ces résultats tendent donc à suggérer l'existence d'une différence de traitement liée au genre en matière de scolarisation et de mise au travail des enfants. C'est, en particulier, cette hypothèse que nous essayons d'examiner dans cette étude en testant l'existence d'un effet genre "discriminatoire" entre les garçons et les filles. Le but étant d'analyser en fonction du genre les décisions de scolarisation et de mise au travail des enfants dans les ménages et de mettre en évidence les sources des disparités entre filles et garçons dans ces décisions.

En effet, deux sources potentielles de disparités peuvent être considérées. La première attribuable aux caractéristiques observables et la seconde imputable aux inobservables.

Toutefois, certaines études évaluent les disparités de genre par un simple test d'égalité des coefficients associés à une variable d'intérêt (sur l'échantillon des filles et des garçons). Ainsi l'existence de disparités entre les filles et les garçons est admise lorsque cette égalité de coefficients entre les deux sous-échantillons est rejetée. Cette méthode est utilisée, par exemple, par Emerson et Souza (2007) qui testent l'égalité du coefficient de la variable *éducation* du père (et de la mère) sur la scolarisation des filles et des garçons. L'un des limites de cette méthode, c'est que qu'elle ne permet pas de déterminer la part des disparités provenant des facteurs inobservables. Une seconde méthode consiste à tester conjointement

l'égalité des coefficients associés à la variable binaire *sexe* sur plusieurs équations à la fois. Par exemple Koissy-kpein (2012) en utilisant les données sur le travail des enfants au Mali, teste l'égalité du coefficient de la variable binaire *filles* entre trois équations de temps de travail (travaux domestiques, travaux économiques et travaux non économiques). En rejetant l'égalité conjointe des coefficients de la variable *filles* sur ces trois équations, l'auteur conclut l'existence d'une disparité de genre en montrant que les filles et les garçons ne sont pas significativement orientés vers les mêmes types de travaux.

Ces deux méthodes diffèrent donc très sensiblement dans leur approche ; la première examine la disparité en considérant les facteurs déterminants et la seconde se focalise sur le résultat observé sans d'avantage préciser ce qui le détermine. Dans cette étude, nous adoptons une démarche consistant à tester, à la fois, les disparités attribuables aux observables et celles dues aux inobservables par l'utilisation de méthodes de décomposition Oaxaca-Blinder (O-B).

Popularisée par Oaxaca (1973) et Blinder (1973), la méthode O-B a été d'abord très largement utilisée dans l'explication des différences de salaires entre les hommes et les femmes sur le marché du travail. Elle a été également utilisée dans de nombreux domaines d'analyse de « différences » comme les comparaisons milieu rural-milieu urbain. Bien que l'application de cette méthode dans les analyses des performances scolaires entre les filles et les garçons se soit sensiblement développée (Lai, 2007 ; Chuy et Nutulescu, 2012 ; Fortin et al., 2013), son application à l'analyse de l'allocation de temps reste encore très limitée, sinon quasi inexistante. Dans cette étude, nous l'utilisons pour tenter de savoir dans quelle mesure la différence dans les indicateurs scolaires et de travail des enfants est influencée par les différences de caractéristiques (observables et inobservables). La méthodologie vise donc à déterminer quelle part pourrait être considérée comme non expliquée par ces différences de caractéristiques. L'équation du test se présente comme suit :

$$\bar{Y}_A - \bar{Y}_B = (\bar{X}_A - \bar{X}_B)\beta_B + \bar{X}_B(\beta_A - \beta_B) + (\bar{X}_A - \bar{X}_B)(\beta_A - \beta_B) \quad (1)$$

Où \bar{Y}_A et \bar{Y}_B représentent respectivement l'espérance de l'outcome d'intérêt sur le groupe des garçons (**A**) et le groupe des filles (**B**). \bar{X}_A et \bar{X}_B représentent l'espérance des vecteurs de caractéristiques sur chaque groupe. β_A et β_B sont des paramètres à estimer. La relation (1) exprime une différence globale $\bar{Y}_A - \bar{Y}_B$. Cette différence peut se décomposer en trois éléments : $(\bar{X}_A - \bar{X}_B)\beta_B$ quantifie la part due aux différences de caractéristiques entre les deux groupes **A** et **B**. Cette part est qualifiée de disparité due aux différences de dotation ou encore de différence expliquée. Elle mesure la variation moyenne de l'outcome du groupe B

s'il avait les mêmes caractéristiques que le groupe A. Le second élément $\bar{X}_B(\beta_A - \beta_B)$ quantifie la disparité due aux différences de coefficients. Il mesure la variation moyenne de l'outcome du groupe B s'il avait les mêmes coefficients que le groupe A. Ce second élément est généralement assimilé à la discrimination ou à la différence d'origine inobservable. Il constitue le point focal de notre test de disparités

Le troisième élément de l'équation est une interaction de ces deux précédentes différences (différence de caractéristiques et différence de coefficients). La décomposition est alors qualifiée de décomposition à trois niveaux. Ici, la décomposition est effectuée du point de vue du groupe des filles, l'analyse peut être renversée du point de vue de garçons lorsqu'on suppose que c'est du côté des garçons que se trouve la discrimination (positive ou négative)².

1. Données et statistiques descriptives

Les données proviennent d'une enquête à deux passages réalisés par le CERDI en 2009 et 2011 au Mali dans le cadre de l'évaluation d'impact du programme Plateformes Multifonctionnelles. La base de données contient un panel de 201 villages et 2412 ménages observés sur deux périodes. Les données sur l'emploi du temps sont collectées à partir d'un module du questionnaire spécifiquement conçu pour recueillir les informations sur les activités quotidiennes des enfants âgés entre 6 et 17 ans. Le nombre total de modules dans l'enquête étant important, tous les enfants dans le ménage n'ont pas été interrogés. L'emploi du temps a été relevé sur un échantillon d'enfants, tiré de façon aléatoire dans le ménage. En 2009, 1984 enfants âgés de 6 à 15 ans ont été observés. En 2011, (au second passage), les mêmes enfants ont été observés. Toutefois 308 enfants de 2009 n'ont pas pu être observés en 2011 pour des raisons diverses (absences, etc.). Par ailleurs, 212 nouveaux enfants de 6 à 15 ans ont été intégrés au suivi en 2011. Au final ce léger mouvement de l'échantillon se solde par une perte sensible de 96 enfants. Le tableau 1 suivant donne la liste des activités codées dans le questionnaire³.

² Une présentation plus détaillée de la méthodologie de décomposition est présentée en Annexe.

³ Cette liste n'étant pas exhaustive, nous avons aussi laissé la possibilité pour la modalité "Autres activités" accompagnée par un champ permettant à l'enquêteur de donner une description de l'activité en question. Ainsi, après traitement l'activité peut être rattachée à l'un des grands postes d'allocation du temps qui ont été préalablement définis.

0 Aucune occupation	13 Ecole	26 Recherche de termites	39 Trajet retour du champ
1 Agriculture	14 Trajet de l'école	27 Chauffage eau	40 Travaux aux champs
2 Elevage	15 Devoirs à la maison	28 Lessive	41 Vente de condiments
3 Aide à la cuisine	16 Jeux	29 Commerce/petit commerce	42 Maraîchage
4 Aller chercher de l'eau	17 Se reposer	30 Activité de cueillette	43 Pâturage
5 Aller chercher du bois	18 Sommeil	31 Toilette	44 Aller autre moulin pour mouture
6 S'occupe des enfants	19 Travail en apprentissage	32 Détente/distraction	45 Orpaillage
7 Nettoyage de la maison	20 Eglise, mosquée ou autre lieux de culte	33 Activités artisanales	46 Autre (non déclaré)
8 Pilage de céréales	21 Prise de repas	34 Travail de banco	
9 Pêche/chasse	22 Transport de fumure	35 Faire abreuver les animaux	
10 Autre aide/travail dans le ménage	23 Trajet du champ	36 Laver les ustensiles de cuisine	
11 Travail non rémunéré, hors ménage	24 S'occuper des invalides	37 Aller à la PTF pour mouture	
12 Travail rémunéré, hors ménage	25 Trajet de recherche de termites	38 Préparation repas	

Sources : Notre enquête

Les travaux menés par les enfants ont été regroupées selon le SCN93⁴. Selon cette classification, l'emploi du temps est défini selon cinq catégories d'activités : les activités productives *économiques*, les activités productives *non-économiques*, les activités *non-productives* (éducation, loisirs) et les activités dites de *nécessités physiologiques* (repos, repas, sommeil, etc.). Le tableau 2 donne la structure de l'emploi du temps des enfants réparti selon ces grands postes d'allocation.

Tableau 2 : L'emploi du temps moyen de l'enfant (en nombre d'heures)

	Ensemble	Filles	Garçons	<i>P.value ttest différence</i>
Activités non-économiques	2.74	4.01	1.49	(0.0000)
Activités économiques	2.29	1.85	2.72	(0.0000)
Activités scolaires	4.88	4.65	5.09	(0.0001)
Activités de loisirs	2.01	1.83	2.18	(0.0000)
Nécessités physiologiques	11.63	11.20	12.05	(0.0000)
Nombre d'observations	3872	1921	1951	

Estimation pooling 2009 et 2011 ; Effets fixes temporels, écart-type calculés en cluster village, Echantillon des enfants de 6-17 ans ;

Ces chiffres présentés du tableau 2 confirment nos premières intuitions sur l'allocation du temps selon le genre. D'abord, nous constatons une grande différence entre les filles et les garçons dans la répartition des activités. Les filles consacrent plus de 4 heures par jour aux travaux non-économiques tandis que les garçons en consacrent largement moins (1,5 heure par jour). Cette inégalité semble se répercuter sur le temps consacré aux activités scolaires puisque les garçons consacrent à ces activités plus de 5 heures par jour alors que les filles dépassent à peine 4,5 heures. On remarque, par ailleurs, que les filles consacrent un peu moins

⁴ Classification adoptée par l'Organisation International du Travail pour mesurer le travail des enfants (voir OIT, 2008).

de temps aux activités économiques que les garçons (1.85h/j contre 2.72 h/j). Toutes ces différences apparaissent significatives au seuil de 1% (voir tableau 2).

Ces résultats soutiennent ainsi l'idée d'une répartition des tâches fortement marquée selon le genre. Les filles sont essentiellement orientées vers des travaux non-économiques, tandis que les garçons sont significativement plus orientés vers les activités économiques. Cependant, cette disparité nette traduit-elle une préférence inobservable des parents basée sur le genre de l'enfant ? La réponse à cette interrogation nécessite une analyse plus poussée par l'utilisation des techniques économétriques.

2. Estimations économétriques et résultats

Trois tests de décomposition sont réalisés. Le premier concerne l'accès à la scolarisation, le second et le troisième portent sur l'allocation du temps aux activités économiques et aux activités non-économiques. Les résultats de ces tests sont présentés dans le tableau 3 suivant.

		Probabilité de scolarisation		Temps travail économique		Temps travail non économique	
		Coef.	P> z	Coef.	P> z	Coef.	P> z
Garçons		0.7842	0.000	2.6791	0.000	1.5084	0.000
Filles		0.7229	0.000	1.8615	0.000	3.9756	0.000
Différence		0.0613	0.000	0.8177	0.000	-2.4672	0.000
Part expliquée		0.0048	0.269	0.0750	0.171	0.0334	0.186
Part non expliquée		0.0565	0.000	0.7427	0.000	-2.5007	0.000
Contribution des caractéristiques à la part expliquée							
Caractéristiques	enfant	0.0063	0.015	0.0115	0.198	-0.0190	0.152
	chef ménage	-0.0012	0.441	0.0015	0.774	0.0073	0.340
	mère	0.0001	0.909	-0.0018	0.666	0.0013	0.766
	ménage	-0.0001	0.935	-0.0062	0.365	0.0043	0.619
	village	-0.0003	0.910	0.0700	0.084	0.0395	0.066
Contribution des coefficients à la part non expliquée							
Coefficients associés aux caractéristiques	enfant	-0.0944	0.048	-0.0151	0.984	0.5674	0.493
	chef ménage	-0.0363	0.754	-0.5229	0.367	-0.5668	0.393
	mère	0.0379	0.716	-0.3395	0.532	0.8670	0.145
	ménage	-0.0912	0.148	0.1075	0.747	-0.7994	0.020
	village	0.0292	0.004	-0.2842	0.184	-0.1227	0.645
	constante	0.2113	0.002	1.7970	0.013	-2.4462	0.004
Observations		3596		3596		3596	

Comme on pouvait s'y attendre, la différence entre les filles et les garçons dans l'accès à la scolarisation est très fortement significative. On constate que l'espérance de la probabilité de scolarisation pour les garçons est de 78.4% tandis que celle-ci est de 72.2% pour les filles, soit une différence d'un peu plus de 6 points de pourcentage. Cette différence apparaît significative au seuil de 1%. En décomposant cette différence en une partie due aux différences de caractéristiques et une partie due aux inobservables, nous trouvons qu'environ 92% de la différence reste inexpliquée (soit 5.6 de point de pourcentage). Cette part non expliquée reste également significative au seuil de 1% tandis que la part expliquée apparaît non significative (p.value=26.9%).

Pour analyser la contribution des caractéristiques à la différence expliquée et la contribution des coefficients associés à ces caractéristiques à la différence non expliquée, nous regroupons les variables en des grandes catégories qualifiées d'attributs⁵. Nous avons ainsi les attributs des enfants, les attributs des parents, ceux du ménage et ceux du village. Ce regroupement montre que l'essentiel de la part non expliquée provient du coefficient associé à la constante (0.2113). La constante représente ici l'effet d'appartenance à l'un des groupes (en l'occurrence celui des garçons). Le coefficient lié aux attributs de l'enfant (-0.0944), étant la seconde contribution en termes absolus, ce coefficient montre par son signe négatif que les différences de réaction ont tendance à contribuer négativement à la différence inexpliquée c'est-à-dire à diminuer et résorber la différence entre les filles et les garçons. En revanche, les différences de coefficients observées à partir des caractéristiques villages tendent à contribuer positivement à l'écart inobservable i.e. à creuser l'écart d'origine inobservable (0.0292). On constate par ailleurs que les contributions des coefficients associés aux attributs du chef de ménage et de la mère sont de même ordre en termes absolus mais de signe contraire (-0.0363 contre 0.0379). Bien que non significatives, ces contributions suggèrent que la différence inobservable réagit négativement aux attributs du chef de ménage (diminution de l'écart inobservable) et réagit positivement aux attributs de la mère. Ce qui aurait traduit une forme de préférence inobservable des mères pour la scolarisation des garçons que des filles.

S'agissant de la différence dans la demande de travail, nous constatons également une forte disparité entre le groupe des filles et celui des garçons. L'espérance de la participation des garçons aux activités économiques est de 2.67 heures par jour contre 1.86 heure pour les

⁵ Dans ce tableau, le terme *attribut enfant* : regroupe les variables liées aux enfants (âge, rangs de naissance, etc.), chef de ménage regroupe les attributs du chef de ménage (âge, éducation, etc.). Ainsi de suite pour le ménage et pour le village (voir la liste des variables dans le tableau A.1 en Annexe).

filles. En outre l'espérance de la participation des garçons aux activités non-économiques est de l'ordre de 1.5 heure comparativement aux filles pour lesquelles la valeur espérée est 3.97 heures. Ces résultats étaient d'ailleurs largement prévisibles à partir des statistiques du tableau 2 qui montraient déjà une très large différence dans l'allocation du temps entre les filles et les garçons. Les résultats du tableau 4 confortent cette idée d'une forte spécialisation dans les travaux selon le genre. Les garçons se spécialisent dans les travaux économiques et tandis que les filles sont beaucoup plus impliquées dans les activités non-économiques. Même si les différences globales entre les filles et les garçons apparaissent toutes significatives au seuil de 1%, la part de ces différences due aux caractéristiques apparaît insignifiante. L'essentiel des différences traduit donc un gap discriminatoire entre les deux groupes.

Conclusion

Cette étude tentait de prolonger la réflexion et les analyses sur les disparités de genre dans l'allocation du temps des enfants. Elle s'est focalisée sur la comparaison des indicateurs de résultat entre les garçons et les filles en partant des cas de la scolarisation et la participation aux travaux. La méthodologie a été essentiellement fondée sur la décomposition Oaxaca-Blinder. Les résultats de cette décomposition montrent que l'essentiel des différences entre les garçons et les filles provient des facteurs inobservables. L'étude met aussi en évidence une forte polarisation des travaux selon le genre. Les garçons sont significativement plus orientés vers les activités économiques (1 heure de plus que les filles par jour) alors que les filles sont significativement plus orientées vers les activités non économiques (une différence avec les garçons d'environ 2.5 heures par jour).

Cependant, l'inégalité d'accès à la scolarisation entre les filles et les garçons et la forte polarisation dans l'exécution des travaux peuvent aussi trouver une large explication dans l'influence des règles socioculturelles. Plusieurs études sociologiques ont déjà mis en avant le rôle des groupes familiaux dans les choix de scolarisation et d'orientation vers les travaux. Par exemple, Clignet (1994), montre qu'à l'intérieur d'un groupe social donné, les garçons et les filles n'ont pas les mêmes chances de fréquenter l'école ni d'y rester car de nombreux parents africains continuent de considérer que "l'occidentalisation" de leur filles les rend rebelles aux rôles domestiques auxquels elles sont destinées. De plus, cet auteur signale le fait que la dot reçue par le ménage au moment du mariage de la fille peut diminuer en fonction de son niveau d'instruction. Par ailleurs, dans les groupes sociaux où les normes sociales font que les activités domestiques sont strictement dévolues aux femmes, cette répartition des tâches peut

aussi se refléter directement dans l'allocation du temps des enfants. Par exemple, les filles participent à ces travaux non-économiques, d'une part pour aider leur mère et d'autre part comme une forme d'apprentissage à la vie de femme adulte. Alors que les garçons considérés comme futurs chefs de ménage doivent accompagner leur père dans les champs agricoles, garder le bétail, etc. Dans ce schéma sociétal, les garçons seront davantage spécialisés dans les travaux économiques et les filles dans les activités non-économiques. Cette spécialisation s'inscrit ainsi dans une forme de reproduction de structure sociale.

Bibliographie

- Basu, 2001, «Gender and Say: A Model of Household Behavior with Endogenously-determined Balance of Power», Cornell University, *mimeo*.
- Blinder, 1973, «Wage discrimination: Reduced form and structural estimates », *Journal of Human Resources* 8: 436–455.
- Chuy et Nitulescu, 2012, «Pisa 2009 Explaining the Gender Gap in Reading through Reading Engagement and Approaches to Learning», Research Paper , Council of Ministers of Education, Canada
- Clignet R., 1994, « La demande d'éducation : aspects sociologiques », *Afrique contemporaine*, n°172, Numéro spécial « Crises de l'éducation en Afrique », Paris, : 108-118.
- Cotton, J., 1988, «On the decomposition of wage differentials ». *Review of Economics and Statistics* 70: 236–243.
- Emerson et Souza, 2007, «Child Labor, School Attendance, and Intrahousehold Gender Bias in Brazil», *World Bank Economic Review* Volume 21
- Fortin et al., 2013, «leaving boys behind: gender disparities in high academic achievement», *Nber Working Paper Series*, no 19331
- Koissy-Kpein, 2012, «Gender and Competition between Economic or Non-economic Labor and Schooling: Evidence from EPAM Mali», *African Development Review*, Vol. 24, No. 1, 2012, 107–123
- Lai, 2007, «Are Boys Left Behind? The Evolution of Gender Gap in Beijing Middle Schools» Unpublished paper , *New York University*
- Neumark, 1988, « Employers' discriminatory behavior and the estimation of wage discrimination », *Journal of Human Resources* 23: 279–295.
- Oaxaca et Ransom, 1994, « On discrimination and the decomposition of wage differentials » *Journal of Econometrics* 61: 5–21.
- Oaxaca, 1973, « Male–female wage differentials in urban labor markets » *International Economic Review* 14: 693–709.
- OIT, 2008, « Statistiques sur le travail des enfants », Rapport III, 18e Conférence internationale des statisticiens du travail, 24 nov.-5 déc. disponible à l'adresse www.ilo.org/global/What_we_do/Statistics/events/icls/lang--fr/docName--WCMS_099580/index.htm.
- Reimers, 1983, « Labor market discrimination against Hispanic and black men» *Review of Economics and Statistics* 65: 570–579.
- Ridao-Cano, 2000, « Child Labor and Schooling in a Low Income Rural Economy,» *University of Colorado*, *mimeo*.
- Wahba, 2005, « The Influence of Market Wages and Parental History on Child Labor and Schooling in Egypt » *Journal of Population Economics*

Annexe

Méthodologie de décomposition Oaxaca-Blinder

Nous partons de l'équation suivante :

$$Y_G = X_G\beta_G + \epsilon_G \quad E(\epsilon_G) = 0 \quad G \in (A, B) \quad (2)$$

Où Y_G est l'outcome sur le groupe G , X_G les caractéristiques de ce groupe, ϵ_G le terme d'erreur dont l'espérance est nulle. β_G un paramètre à estimer sur les deux groupes A et B d'analyse (soit respectivement les garçons et les filles). La différence entre les garçons et les filles peut simplement se traduire à partir de cette équation :

$$D = E(Y_A) - E(Y_B) = E(X_A)\beta_A - E(X_B)\beta_B \quad (3)$$

Pour identifier la contribution des caractéristiques à cette différence, l'équation (3) peut se réécrire selon la décomposition O-B comme suit :

$$D = \{E(X_A) - E(X_B)\}\beta_B + E(X_B)(\beta_A - \beta_B) + \{E(X_A) - E(X_B)\}(\beta_A - \beta_B) \quad (4)$$

En posant $E(X_A) = \bar{X}_A$ et $E(X_B) = \bar{X}_B$ où \bar{X}_A et \bar{X}_B représentent l'espérance ou les valeurs moyennes de la matrice des variables de prédiction, on a :

$$D = (\bar{X}_A - \bar{X}_B)\beta_B + \bar{X}_B(\beta_A - \beta_B) + (\bar{X}_A - \bar{X}_B)(\beta_A - \beta_B) \quad (5)$$

Cette différence est constituée de trois éléments :

$(\bar{X}_A - \bar{X}_B)\beta_B$ quantifie la part de la différence due aux différences de caractéristiques entre les deux groupes. Elle est qualifiée de disparité due aux différences de dotation ou encore de différence expliquée. Elle mesure la variation moyenne de l'outcome du groupe B s'il avait les mêmes caractéristiques que A. Le second élément $\bar{X}_B(\beta_A - \beta_B)$ quantifie la disparité due aux différences de coefficients. Il mesure la variation moyenne de l'outcome du groupe B s'il avait les mêmes coefficients que le groupe A. Le troisième élément est une interaction des deux différences (différence de caractéristiques et différence de coefficients). La décomposition est alors dite décomposition à trois niveaux.

Ici, la décomposition est effectuée du point de vue du groupe des filles, l'analyse aussi peut être renversée du point de vue de garçons lorsqu'on suppose que c'est du point de vue des garçons que pourrait se trouver la discrimination. A cet égard, de nombreux auteurs

recommandent l'utilisation d'un coefficient neutre pour analyser les différences. Soit β^* un tel coefficient, alors la différence exprimée par l'équation (5) devient :

$$D = (\bar{X}_A - \bar{X}_B)\beta^* + \bar{X}_A(\beta_A - \beta^*) + \bar{X}_B(\beta^* - \beta_B) \quad (6)$$

Ainsi en choisissant un coefficient de référence β^* on obtient une décomposition à deux niveaux où le premier niveau $(\bar{X}_A - \bar{X}_B)\beta^*$ capte la différence de caractéristiques et le second $\bar{X}_A(\beta_A - \beta^*) + \bar{X}_B(\beta^* - \beta_B)$ la partie non expliquée. Cette partie non expliquée est généralement attribuée à la discrimination mais aussi à des caractéristiques inobservables.

La partie non expliquée peut aussi être décomposée en supposant par exemple que :

$$\beta_A = \beta^* + \delta_A \text{ et } \beta_B = \beta^* + \delta_B$$

Où δ_A et δ_B représentent des caractéristiques inobservables ou discrimination liées spécifiquement aux deux groupes (positive ou négative selon le signe). Par exemple au lieu de considérer que ce sont les filles qui sont défavorisées, il est possible que ce soit les garçons qui soient favorisés. Ainsi de façon générale, la différence traduite par l'équation (6) se réécrit comme :

$$D = (\bar{X}_A - \bar{X}_B)\beta^* + \bar{X}_A\delta_A - \bar{X}_B\delta_B \quad (7)$$

Ainsi la partie non expliquée $\bar{X}_A\delta_A - \bar{X}_B\delta_B$ peut donc se diviser en une partie en faveur des garçons $\bar{X}_A\delta_A$ et une partie considérée en défaveur des filles $-\bar{X}_B\delta_B$.

Quant à la détermination de la partie expliquée et la partie non expliquée, elle est faite de manière directe dans la décomposition à trois niveaux, puisque les coefficients β_A et β_B peuvent être directement obtenus à partir d'une régression sur les deux sous échantillons par une méthode d'estimation appropriée (MCO ou probit selon la nature de la variable). Dans ce cas la différence s'obtient simplement par :

$$D = (\bar{X}_A - \bar{X}_B)\hat{\beta}_B + \bar{X}_B(\hat{\beta}_A - \hat{\beta}_B) + (\bar{X}_A - \bar{X}_B)(\hat{\beta}_A - \hat{\beta}_B) \quad (8)$$

Où $\hat{\beta}_A$ et $\hat{\beta}_B$ représentent les valeurs estimées des coefficients sur les deux groupes A et B.

Mais dans une décomposition à deux niveaux, notamment pour l'estimation de β^* , plusieurs suggestions ont été faites dans la littérature compte tenu de la difficulté de choisir le groupe de référence. Par exemple supposer que $\beta^* = \beta_A$ si ce sont les filles qui sont discriminées ou bien supposer $\beta^* = \beta_B$ si ce sont les garçons qui sont discriminés. Oaxaca (1973) à parlé à cet

égard de "*Index number problem*". Certains auteurs comme Reimers (1983) propose d'utiliser une moyenne pondérée des deux coefficients c'est-à-dire de prendre : $\hat{\beta}^* = 0.5\hat{\beta}_A + 0.5\hat{\beta}_B$. Alors que Cotton (1988) suggère plutôt d'utiliser : $\hat{\beta}^* = \left(\frac{n_A}{n_A+n_B}\right) \cdot \hat{\beta}_A + \left(\frac{n_B}{n_A+n_B}\right) \cdot \hat{\beta}_B$ où n_A et n_B représentent respectivement le nombre d'individus dans le groupe A et le groupe B. Tandis que Neumark (1988) défend l'utilisation d'un coefficient estimé par pooling sur l'échantillon total pour dériver $\hat{\beta}^*$. Finalement Oaxaca et Ransom (1994) proposent une forme générale de pondération Ω telle que :

$$D = (\bar{X}_A - \bar{X}_B)[\Omega\beta_A + (I - \Omega)\beta_B] + [(I - \Omega)' \bar{X}_A + \Omega' \bar{X}_B](\beta_A - \beta_B) \quad (9)$$

Où Ω est la matrice de poids relatif sachant les coefficients du groupe A. I représente la matrice identité. La généralité de cette forme de pondération permet par exemple de retrouver $\beta^* = \beta_A$ lorsque $\Omega = I$. Elle permet également de retrouver $\hat{\beta}^* = 0.5\hat{\beta}_A + 0.5\hat{\beta}_B$ lorsque $\Omega = 0.5I$. Oaxaca et Ransom(1994) montre en outre qu'en calculant : $\Omega = (X_A X'_A + X_B X'_B)^{-1} X_A X'_A$ avec les données actuelles observées équivaldrait à l'estimation en pooling de Neumark (1988).

C'est donc cette méthode de pondération que nous mettons en œuvre en estimant les coefficients de référence sur les données groupées. Nous effectuons la décomposition sur trois indicateurs, la probabilité de scolarisation, le temps d'activités économiques et le temps d'activités non économiques.

En retenant ces trois indicateurs, la contribution de chaque caractéristique à la différence expliquée (DE) est simplement déterminée à travers la formule suivante :

$$DE = (\bar{X}_A - \bar{X}_B) \hat{\beta}_A = \sum_{j=1}^k (\bar{x}_{jA} - \bar{x}_{jB}) \hat{\beta}_{jA} \quad (10)$$

Où DE est la différence totale expliquée et \bar{x}_{jA} et \bar{x}_{jB} la moyenne de la j ième variable explicative sur le groupe A et B. et $\hat{\beta}_{jA}$ est le coefficient estimé sur cette variable à partir du groupe A. Donc $(\bar{x}_{jA} - \bar{x}_{jB}) \hat{\beta}_{jA}$ représente la contribution de cette variable à la différence. Pareillement la contribution des coefficients à la différence totale non expliquée (DNE) s'obtient par:

$$DNE = \bar{X}_B (\hat{\beta}_A - \hat{\beta}_B) = \sum_{j=1}^k \bar{x}_{jB} (\hat{\beta}_{jA} - \hat{\beta}_{jB}) \quad (11)$$

Tableau A.1. récapitulatif et description des variables d'analyse

Variables	Définition et mode de calcul	Obs	Moyenne	Ecart-type	Min	Max
scolarisation	1 si l'enfant est scolarisé 0 sinon	3872	0,713	0,448	0	1
travail_non_eco	1 si l'enfant participe aux travaux non économiques 0 sinon	3872	0,487	0,361	0	1
travail_eco	1 si l'enfant participe aux travaux économiques 0 sinon	3872	0,291	0,498	0	1
travail_product	1 si l'enfant participe à l'une des deux types d'activités et 0 sinon	3872	0,672	0,399	0	1
indice_biens_eq	Indice de richesse obtenu par ACP sur les biens et actifs durables (1 ^{ère} composante)	3872	-0,0051	1,677	-5,03	8,96
indice_logement	Indice de richesse obtenu par ACP sur les biens et actifs durables (2 ^{ème} composante)	3872	0,0181	1,388	-6,93	5,38
surface_total_terre	Surface totale en hectares des terres agricoles du ménage	3872	7,664	9,456	0	47,3
education_chef_men	1 si le chef ménage est alphabétisé et 0 sinon	3778	0,327	0,469	0	1
age_chef_men	Age en années révolues du chef de ménage	3827	49,29	11,88	36	82
education_mere	1 si la mère de l'enfant est alphabétisée et 0 sinon	3715	0,124	0,329	0	1
age_mere	Age en années révolues de la mère de l'enfant	3670	36,85	9,268	15	64
taille_menage	Nombre d'individus dans le ménage	3872	8,376	3,265	3	30
menage_polygame	1 si l'enfant vit dans un ménage polygame, 0 sinon	3737	0,379	0,485	0	1
age_enfant	Age en années révolues de l'enfant	3872	11,12	2,902	6	17
sexe_enfant	Sexe de l'enfant 0 féminin 1 masculin	3872	0,504	0,500	0	1
rang_naissance1_men	Rang de naissance de l'enfant parmi tous les enfants du ménage	3710	2,712	1,827	1	12
rang_naissance2_mere	Rang de naissance de l'enfant parmi les enfants de sa mère	3752	2,318	1,491	1	8
lien_biologique_chef_men	1 si l'enfant considéré est le fils du chef de ménage, 0 sinon	3855	0,936	0,244	0	1
distance_primaire	Distance en km du village à l'école primaire la plus proche, 0 si dans le village	3872	0,124	4,635	0	25
distance_college	Distance en km du village au collège le plus proche, 0 si dans le village	3857	2,220	5,815	0	80
distance_lycee	Distance en km du village au lycée le plus proche, 0 si dans le village	3820	41,67	50,64	0	300
frais_scolaires_moyen_primaire	Montant en francs CFA des frais scolaires (Out-of-Pocket) moyen par élève du primaire calculé au niveau du village	3872	11448	5536	1250	35893
frais_scolaires_moyen_college	Montant en francs CFA des frais scolaires (Out-of-Pocket) moyen par élève du collège calculé au niveau du village	3872	23820	17074	4500	150000
frais_scolaires_moyen_lycee	Montant en francs CFA des frais scolaires (Out-of-Pocket) moyen par élève du lycée calculé au niveau du village	3596	58183	44642	5000	350000
nbre_habitants_village	Nombre d'habitants dans le village	3843	1818	1745	102	11271
distance_route	Distance en km du village à la route bitumée ou la piste la plus proche, 0 si passet dans le village	3872	3,913	13,73	0	100
distance_marche	Distance en km du village au marché le plus proche, 0 si dans le village	3781	6,676	9,922	0	90