

# AFRILUX

## Impact des capitaux économique et culturel des familles sur la scolarisation des enfants au Mali

Marie BOLTZ, Ronan LE SAOUT, Paul PIVETEAU, Gabriel SMAGGHUE  
sous la direction de Mathias KUEPIE - CEPS/INSTEAD



### SOMMAIRE

I. SYNTHÈSE .....	1
II. INTRODUCTION .....	4
III. CADRAGE CONTEXTUEL ET THEORIQUE .....	5
1. Contexte éducatif, social et économique au Mali .....	5
2. Revue de littérature .....	6
IV. ANALYSES DESCRIPTIVES .....	8
1. Descriptif des données .....	8
2. Le capital éducatif .....	9
3. Le capital économique .....	10
4. La fréquentation scolaire selon nos deux dimensions .....	11
V. MODELISATION ECONOMETRIQUE .....	12
1. Cadre théorique .....	12
2. Pondérer ou non le modèle ? .....	13
3. Quelle spécification pour les variables d'intérêt .....	13
4. Résultats du modèle simple .....	14
5. Tenir compte du plan de sondage pour corriger l'hétéroscédasticité .....	15
6. L'endogénéité des variables .....	16
7. Valeurs aberrantes et qualité d'ajustement du modèle .....	18
VI. INTERPRETATIONS DES RESULTATS .....	19
1. Analyse des effets marginaux .....	19
2. Etude de la stabilité des paramètres entre genres .....	20
CONCLUSION .....	22
BIBLIOGRAPHIE .....	23
ANNEXES .....	24

### I. Synthèse

Le Mali fait partie des pays les moins avancés (PMA). Il est classé 175ème sur 177 en 2004 selon l'indicateur de Développement Humain. En outre, comme d'autres pays d'Afrique subsaharienne, le Mali présente des taux de scolarisation très faibles. Ainsi, seulement 47% des enfants en âge d'être scolarisés (de 7 à 15 ans) l'étaient vraiment en 2003. Autre indicateur clé, le taux d'alphabétisation pour les 15 ans et plus est seulement de 26%.

Face à de tels chiffres, le gouvernement malien, dans le cadre du Programme Décennal de Développement (2000-2010), s'est fixé pour objectif d'atteindre d'ici 2010 un taux de scolarisation de 95% pour le premier cycle de l'enseignement fondamental (équivalent du primaire en France). L'offre d'éducation a progressé. Il est également pertinent de se poser la question de savoir qui va à l'école. Quel est le poids des capitaux économique (i.e. leur niveau de vie) et culturel (mesuré par le "stock" d'éducation des adultes) des ménages dans la scolarisation de leurs enfants ?

Dans cette étude, nous étudions donc l'influence, sur la probabilité de fréquentation scolaire des enfants, de ces deux dimensions caractéristiques majeures des ménages. Pour ce faire, nous nous appuy-

ons sur des données provenant de l'Enquête Légère Intégrée auprès des Ménages (ELIM), réalisée de juillet à novembre 2006 au Mali. Cette enquête a été menée par sondage auprès d'un échantillon de 4494 ménages représentatifs de l'ensemble du pays.

Afin d'appréhender les spécificités du cas Malien et d'identifier notre population, nous avons étudié l'organisation du système éducatif du pays. D'autre part, une revue de littérature sur les déterminants de l'éducation dans les pays en développement a permis de mettre en évidence les enjeux théoriques afférents à notre étude. Les analyses descriptives conduites permettent quant à elles l'identification des particularités du Mali et des variables pertinentes pour la suite de notre étude. Enfin, l'analyse économétrique menée à partir de régressions sur variables dépendantes dichotomiques (régression logistique ou probit) a permis de mesurer les effets de nos dimensions contrôlés de l'influence d'autres variables. Elle a donné lieu à de nombreuses réflexions techniques. Les analyses ont été effectuées à l'aide du logiciel STATA.

### Cadre contextuel : le choix de notre population d'intérêt, les enfants de 6 à 15 ans

Le système éducatif malien est organisé en deux cycles d'enseignement

fondamental de six et trois années correspondant à l'école primaire et au collège en France. Le lycée quant à lui dure trois années comme en France. Les enfants de 6 à 15 ans sont donc notre population d'intérêt. Nous avons fait le choix d'inclure les enfants de six ans car un nombre non négligeable d'enfants sont déjà scolarisés à cet âge. Les 15 ans représentent l'année de l'entrée sur le marché du travail. Nous avons traité ces âges extrêmes de manière particulière dans notre étude.

### **Cadre théorique : l'éducation comme un coût d'opportunité mais des externalités positives liées à l'éducation**

De prime abord, les travaux de Gary Becker sur la théorie du capital humain et le rôle de la famille nous intéressent tout particulièrement pour leur analyse du choix par les parents d'investissement en capital humain pour les enfants. Cette décision est caractéristique d'un arbitrage coûts-bénéfices.

Concernant l'impact du **capital économique**, la scolarisation génère, en plus des coûts directs (uniformes, matériels scolaires...), des coûts d'opportunité sur le revenu du ménage. En effet, lorsqu'il va à l'école, l'enfant coûte à ses parents la valeur de l'activité productive qu'il ne réalise pas.

De récentes études expérimentales ou liées à des avancées institutionnelles, par exemple le fait de mettre en place la gratuité de l'école au Kenya ou en Ouganda, montre que la réduction des coûts permet d'augmenter les inscriptions et de réduire les abandons. Quant aux bénéfices de l'école, ils sont de deux ordres: les rendements économiques attendus de la formation, et le prestige que les parents retirent d'enfants bien instruits.

Ceci nous souligne l'importance du capital éducatif du ménage. En effet, la plus ou moins bonne perception de ce bénéfice dépend de la connaissance qu'ont les parents du système scolaire, donc de leur propre niveau d'instruction. Or, dans les pays comme le Mali, la très large

majorité des parents ne sait ni lire ni écrire. Il apparaît donc intéressant de mesurer l'impact du **capital éducatif** du ménage sur la scolarisation des enfants. Les efforts d'information sur les rendements de l'éducation (à Madagascar ou en République Dominicaine) ont permis de montrer l'importance du capital éducatif, en tant que connaissance du système éducatif, dans le processus de décision des ménages.

Afin de tenir compte de la perturbation possible de la relation entre fréquentation de l'école et les deux dimensions explicatives principales (capital économique et stock d'éducation), un certain nombre de variables de contrôle (notamment les caractéristiques démographiques des familles et des enfants, le milieu de résidence, le genre etc.) doivent également être introduites dans les modèles.

### **Analyses descriptives : fréquenter le premier cycle fondamental n'assure pas de savoir lire et écrire**

L'étude descriptive de nos variables d'intérêt nous a permis de sélectionner les indicateurs associés au capital économique et éducatif. Le capital économique est mesuré par la dépense totale du ménage par tête car cette variable est la plus discriminante. Concernant le capital éducatif, avoir validé les cinq premières années d'étude n'implique pas de savoir lire et écrire. Ainsi, le capital éducatif est mesuré par deux variables, le niveau d'étude maximal du ménage (pour les individus de plus de 15 ans) et la capacité à lire et à écrire du chef de ménage. Le fait de retenir le niveau d'étude maximal au sein du ménage et non celui du seul chef de famille provient du fait qu'il peut exister des effets d'entraînement des conjoints, des frères et soeurs, ou d'autres membres du ménage. Les taux de scolarisation sont fortement discriminés par modalités. Pour le capital éducatif, ils s'étendent de 30% lorsqu'aucun adulte du ménage n'a reçu d'éducation à 80% lorsque ce niveau est au moins le lycée et plus, en passant par 50% et 65% pour les deux premiers cycles de

l'école fondamentale. Pour le capital économique, l'étendue est moins importante, les résultats par quartiles croissants de la consommation par tête étant respectivement 35%, 44%, 56% et 76%. Par ailleurs, on constate une relation non linéaire entre nos deux dimensions d'étude et la probabilité de scolarisation au sein d'un ménage.

### **Modélisation économétrique : spécification et diagnostic, correction de différents biais potentiels**

Nous étudions les effets spécifiques des deux dimensions de notre analyse, à savoir le capital économique et éducatif du ménage, sur la scolarisation des enfants au Mali. Cependant, d'autres facteurs affectent également la décision de scolarisation de l'enfant. Nous nous employons donc à identifier ces facteurs et les utiliser comme variables de contrôle dans notre analyse. Pour ce faire, nous disposons, de par l'enquête ELIM, de multiples variables qui nous permettent de contrôler l'aspect offre d'éducation, les inégalités en termes de situations géographiques, d'accessibilité aux infrastructures et d'autres contraintes extérieures pesant sur les ménages dans leur choix de scolariser leurs enfants.

Un modèle non linéaire est ainsi particulièrement adapté à la nature qualitative et dichotomique de notre variable à expliquer, soit la fréquentation ou non de l'école par l'enfant. Pour des raisons à la fois statistique et pratique, nous avons fait le choix de ne pas pondérer les données. Par ailleurs, nous avons opté pour une partition en classes de nos variables explicatives d'intérêt. Un modèle avec une spécification en continu de ces variables a néanmoins permis de juger non significatif l'effet croisé entre capital éducatif et capital économique. Un premier modèle est construit, mais dont l'ajustement n'est pas significatif. Il a donc été amélioré. Tout d'abord, l'hypothèse d'homoscédasticité a peu de chance d'être vérifiée puisqu'au sein d'un ménage, les enfants partagent certainement une culture familiale ou d'autres caractéristiques non observ-

ables et expliquant la scolarisation. Ensuite, d'un point de vue théorique et notamment d'après les théories de la fécondité endogène (Becker), des biais d'endogénéité méritent d'être testés. Le nombre d'enfants au sein d'un ménage n'apparaît pas endogène, le travail des enfants si. Enfin, les valeurs aberrantes et fortement influentes ont été retirées de notre modèle.

Nous avons donc dû complexifier notre modèle initial, pour obtenir un modèle valide pour l'interprétation. On retrouve le même type d'analyse que pour le modèle simple : les effets du capital économique et éducatif sont croissants et significatifs. Par contre, le fait de savoir lire et écrire pour le chef de ménage n'est pas significatif et n'apparaît donc pas comme un critère pertinent pour la mesure du capital éducatif. Plusieurs remarques sont importantes. Tout d'abord, tenir compte de la structure du plan de sondage est indispensable, certains résultats de tests s'en trouvent en effet modifiés. Cela demeure néanmoins complexe. De plus, nous ne sommes pas à l'abri d'un biais de variables omises ou d'erreurs d'interprétation des variables (liées à la compréhension du questionnaire). En effet, le contrôle de l'offre d'éducation (à travers la distance à l'école) est faible, on ne tient pas compte des types d'école accessibles, de la taille des classes (saturation). Les dépenses liées à l'éducation, la langue parlée ne sont pas prises en compte directement. Enfin, le traitement du caractère endogène du travail des enfants mériterait d'être amélioré, la correction ici proposée étant biaisée.

### **Interprétation des résultats : le capital éducatif a plus de poids que le capital économique**

Dans un modèle non linéaire, les signes des paramètres estimés nous informent du sens de variation de la probabilité de scolarisation. Par contre, à la différence des modèles linéaires, les effets marginaux (i.e. la variation de la probabilité de scolarisation estimée lorsque l'on fait varier une variable explicative d'une unité) ne sont pas constants car ils dépendent de la valeur de tous les coefficients et de toutes

les variables, en d'autres termes de la position où l'on se situe pour les évaluer. Pour pallier ce problème, nous calculons la moyenne des effets marginaux individuels. L'analyse de ces effets marginaux de notre modèle économétrique indique alors qu'un accroissement des capitaux économique et éducatif du ménage permet une fréquentation scolaire accrue des enfants. En outre, le capital éducatif a un effet prépondérant au regard de celui du capital économique. De plus, les chances de scolarisation d'un enfant sont d'autant plus grandes que le niveau de consommation par tête total du ménage est important. A l'inverse, on observe concernant le capital éducatif maximal du ménage un phénomène de saturation pour les niveaux d'éducation supérieurs.

Par ailleurs, de nombreuses inégalités dans la scolarisation des enfants au Mali sont constatées que ce soit en termes de milieu (urbain, rural) ou de genre (fille, garçon). Plus particulièrement, un argument souvent avancé dans la littérature suggère l'existence de différences dans les déterminants de la fréquentation scolaire d'un enfant, selon que celui-ci soit un garçon, ou une fille. Nous nous sommes donc intéressés aux manières de spécifier cette caractéristique du genre. Notre modèle initial se contentait d'introduire une indicatrice pour capter l'effet engendré par le fait d'être une fille. Ainsi, dans la version classique de notre modèle, cet effet était négatif et significatif, traduisant la fréquentation plus faible des filles du système scolaire. Dans un second temps, on teste la pertinence d'un modèle opérant une séparation plus marquée, i.e. qui consiste à séparer totalement les deux échantillons, en partant de l'hypothèse que les ressorts de l'éducation ne sont pas les mêmes entre les deux sexes. Le résultat en est que l'introduction d'une indicatrice est suffisante pour prendre en compte les effets du sexe sur le rôle des capitaux économique et éducatif. L'influence des deux dimensions de notre modèle est donc statistiquement non différente selon le genre de l'enfant.

### **Conclusion**

Ainsi, la fréquentation scolaire des enfants en âge d'être scolarisés au Mali est largement déterminée par le milieu familial d'appartenance, plus précisément par le pouvoir d'achat du ménage et sa connaissance (et reconnaissance) du système éducatif malien. Un enseignement certain de notre étude est que l'influence du capital éducatif est largement prépondérante sur celle du capital économique dans la décision de scolarisation des enfants. L'implication de ce résultat en termes de politiques économiques serait une concentration des pouvoirs publics sur l'offre d'éducation, l'alphabétisation des adultes, des campagnes de sensibilisation et d'informations quant au système scolaire malien, bien plus que sur la mise en place de transferts monétaires vers les ménages. Ce résultat est d'autant plus intéressant que dans un contexte de pays très pauvre où les pouvoirs publics ont des moyens limités, ces recommandations apparaissent moins coûteuses.

## II. Introduction

L'éducation constitue, de nos jours, un des facteurs majeurs de productivité, de développement économique et de progrès social (Schultz, 1961 ; Becker, 1993). Elle a notamment joué un rôle déterminant dans l'essor des pays qui ont pu, malgré un important déficit de ressources naturelles, initier un processus de développement économique soutenu (Japon, Corée, etc.). A l'inverse, l'absence ou la faible accumulation du capital humain constitue une cause majeure du manque de dynamisme économique d'un certain nombre de pays pourtant dotés d'énormes richesses naturelles. Sur le plan sociodémographique, la faiblesse de la scolarisation explique les niveaux préoccupants des indicateurs de santé (espérance de vie faible, taux de mortalité et de fécondité élevés).

Le Mali, comme un certain nombre de pays africains, présente des niveaux de scolarisation très faibles. Malgré d'importants progrès accomplis dans ce domaine les vingt dernières années, l'objectif de scolarisation universelle est loin d'être atteint. Les niveaux d'éducation moyen de la population et d'alphabétisation des adultes demeurent très faibles. Ainsi, sur la tranche d'âge de scolarité obligatoire (7-15 ans), seulement 47% des enfants étaient effectivement scolarisés en 2003. De nombreux facteurs, tant du côté de l'offre que de celui de la demande en éducation co-agissent dans la décision de scolarisation d'un enfant.

L'objectif de cette étude est l'analyse des déterminants de la demande en éducation primaire des ménages au Mali. Plus précisément, nous étudions l'influence, sur les chances de fréquentation scolaire des enfants, de deux dimensions caractéristiques majeures des ménages : leur capital économique (mesuré par les dépenses totales de consommation) et leur capital culturel (mesuré par le "stock" d'éducation des adultes du ménage).

Pour ce faire, nous avons utilisé des données provenant de

l'Enquête Légère Intégrée auprès des Ménages (ELIM), réalisée de juillet à novembre 2006 au Mali. L'ELIM est une enquête par sondage menée auprès d'un échantillon de 4494 ménages représentatifs de l'ensemble du pays, soit une population de 40810 individus.

La suite de cette étude est organisée comme suit. Tout d'abord, dans une perspective générale, l'étude du contexte éducatif, social et économique du Mali est indispensable à une bonne compréhension de notre population d'intérêt. La scolarité obligatoire court de 7 à 15 ans ; ce qui correspond aux deux cycles de l'école fondamentale, ou bien à titre comparatif, en France, à l'école primaire et au collège. Des inégalités en termes de fréquentation de l'école entre genre et milieu (rural ou urbain) sont constatées. Nous nous intéressons dans notre étude aux enfants de 6 à 15 ans. Cependant, les âges extrêmes rendent compte de phénomènes spécifiques et seront traités séparément: une scolarisation précoce pour les enfants de 6 ans, l'âge légal du début de la vie active pour les enfants de 15 ans. Concernant les caractéristiques démographiques, le Mali est un pays faiblement urbanisé à population majoritairement jeune.

Dans la littérature, deux types d'études se distinguent sur ces questions. La première s'appuie sur les travaux de Gary Becker sur la théorie du capital humain et le rôle de la famille. Dans ce cadre, le choix d'investissement en capital humain, par les parents ou les enfants, est fonction d'une analyse des coûts et des bénéfices associés. Les suivantes, plus récentes et empiriques, visent à évaluer les effets de politiques publiques ou programmes particuliers dans le domaine de l'éducation dans les pays en développement. Ces études montrent que les capitaux économique et culturel ont un impact positif sur la scolarisation des enfants. Elles identifient également de nombreux autres facteurs, tels que les caractéristiques démographiques des familles et des enfants, le milieu de résidence, l'offre d'éducation, qui

formeront les variables de contrôle des modèles étudiés.

Les analyses descriptives constituent une première étape dans la caractérisation des corrélations entre les capitaux économique, éducatif et la fréquentation scolaire. Les indicateurs à prendre en compte sont ainsi identifiés : les dépenses totales de consommation par membres du ménage pour la mesure du capital économique, le niveau d'éducation maximal des membres du ménage associé au fait de savoir lire et écrire pour la mesure du capital éducatif. En effet, il est mis en évidence que les cinq premières années d'étude ont des effets ambigus: elles n'assurent pas de savoir lire et écrire, ni d'avoir un impact positif sur les dépenses de consommation.

Enfin, un modèle économétrique est mis en place à partir de techniques de régressions sur variables dépendantes dichotomiques (régression logistique). Des variables de contrôles, telles que les caractéristiques démographiques des familles et des enfants, ou encore le milieu de résidence, ont été introduites dans les modèles. Ceci permet d'éviter des perturbations extérieures de la relation entre la variable à expliquer et les deux dimensions explicatives principales. Un modèle de base est ainsi construit, mais dont l'ajustement n'est pas adéquat. Ce modèle est ensuite amélioré pour tenir compte des variables endogènes, des données aberrantes, de la structure des données (le plan de sondage au niveau des ménages) et corriger la matrice de variance-covariance. Les tests de qualité d'ajustement du modèle ont également été corrigés, pour tenir compte du plan de sondage. L'interprétation des résultats se concentre sur les effets marginaux et une analyse de la stabilité des paramètres par genre. Les analyses sont effectuées à l'aide du logiciel STATA (cf. figure 1)

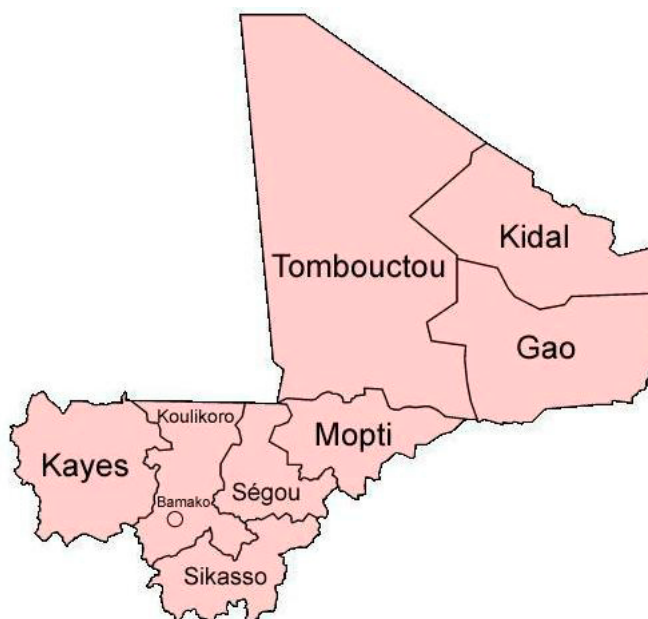


### 1. Contexte éducatif, social et économique du Mali

Le Mali appartient au groupe des pays les moins avancés au monde. En prenant pour indicateur l'IDH, il est classé 175<sup>ème</sup> pays sur 177 en 2004<sup>1</sup>. Autre indicateur notoire, l'espérance de vie à la naissance au Mali est de 49 ans en 2005. Sur le plan économique, le PIB malien par habitant est de \$1200 (en Parité de Pouvoir d'Achat) en 2008 selon le CIA World Factbook. 3.8 % de la surface du pays est considérée comme arable. Le pays est faiblement urbanisé : 31,7% des habitants vivent en milieu urbain. A titre de comparaison, en 2005 le taux d'urbanisation en France est de 77%. La région fluviale autour du fleuve Niger concentre la majorité de la population et les principales activités économiques. Près de 80% de la population active au Mali travaille dans l'agriculture ou la pêche. L'activité industrielle est également concentrée autour des activités agricoles. Le pays est donc très vulnérable aux fluctuations des prix mondiaux de matières premières notamment le coton, son exportation principale. Les transferts financiers des immigrants des pays développés constituent une très importante manne de rêve.

#### Le système éducatif malien et la fréquentation scolaire

Le système éducatif malien est hérité des réformes de 1962, après l'avènement de l'indépendance. L'éducation de base est encore loin d'être universelle. Ainsi, les taux nets de scolarisation<sup>2</sup> de l'école primaire (premier cycle de l'école fondamentale) et du collège (seconde cycle de l'école fondamentale) sont de 55% et 17% en 2006, d'après les chiffres issus de l'enquête ELIM. Les taux bruts sont respectivement de 74% et 37%. Le taux de fréquentation scolaire (quel que soit le niveau) des enfants de 7 à 14 ans est de 53,9%. Il est plus faible pour les enfants de



Source : Wikipedia

6 ans (18,1%) et correspond à une scolarité précoce. Il est également plus faible pour les enfants de 15 ans (42,1%) et correspond à l'âge légal de début de la vie active. Le taux d'alphabétisation pour les 15 ans et plus est de 26%.

Face à de tels chiffres, il est important de comprendre le fonctionnement de ce système afin de saisir les bouleversements de l'offre d'éducation que connaît actuellement le Mali. Le pays s'est fixé l'éducation comme une priorité depuis l'indépendance en 1960. Sous le régime autoritaire des années 1970 et 1980, cet objectif a largement été mis entre parenthèse et les effectifs scolaires ont connu un déclin certain. De plus, au moment de leur indépendance, les pays d'Afrique francophone ont choisi la formation d'une élite. Au contraire, les pays anglophones ont privilégié l'éducation de base. Cette situation explique des taux de scolarisation primaire globalement plus faibles dans les pays d'Afrique francophone (Lange, 1999). Depuis l'instauration de la Troisième République (1991), l'éducation réintègre son rôle de

premier plan dans les priorités étatiques. En 1999, a été élaboré un Programme décennal de développement de l'éducation (PRODEC 2000-2010). Ce programme fixe comme objectif d'atteindre d'ici 2010 un taux brut de scolarisation de 95% pour le premier cycle de l'enseignement fondamental. Définissant onze axes prioritaires, le PRODEC met l'accent sur la volonté politique d'offrir une éducation de base de qualité pour tous.

Le système éducatif malien se décompose en l'école fondamentale, le lycée et les écoles professionnelles. L'enseignement pré-primaire reste encore peu développé. L'école fondamentale dure neuf ans dont six pour le premier cycle (équivalent du primaire et sanctionné par le Certificat de fin d'études du premier cycle de l'enseignement fondamental) et trois années pour le second cycle (équivalent du collège et sanctionné par le Diplôme d'études fondamentales). Le lycée dure quant à lui 3 ans, les écoles professionnelles de 2 à 4 ans. La scolarité obligatoire est d'une durée de 7 ans (7-15 ans).

<sup>1</sup> Sauf indication contraire, les statistiques de cette partie sont issues de l'enquête ELIM 2006 (décrites dans la partie suivante) ou d'organismes internationaux (UNESCO, UNICEF...).

<sup>2</sup> Le taux net de scolarisation correspond au ratio des effectifs d'un groupe d'âge officiel dans un degré donné d'enseignement sur la population correspondante. Le taux brut prend en compte au numérateur l'ensemble des effectifs dans un degré spécifique d'enseignement, quel que soit leur âge.

Face à des contraintes de moyens, l'Etat malien reconnaît en 1994 l'enseignement privé, tandis que les écoles communautaires créées, gérées et financées par des communautés villageoises ou des associations bénéficient de la reconnaissance et du soutien de l'Etat. Outre ces écoles publiques, privées et communautaires, il existe au Mali les " médersas ", écoles dans lesquelles le même programme d'enseignement fondamental est dispensé mais en langue arabe. Si ces écoles ne sont implantées au Mali que depuis une quinzaine d'années, elles connaissent un développement considérable. 77,1% élèves de 6 à 15 ans fréquentant l'école sont inscrits dans une école publique de l'état, 12,5% dans une école communautaire et 10,5% dans une école privée.

Les inégalités (sexuelle, ethnique, géographique) en termes de fréquentation scolaire sont encore très présentes. Les inégalités entre sexes existent à tous les niveaux du système éducatif, mais elles sont d'autant plus grandes que le niveau scolaire atteint est haut. Elles s'expliquent par de nombreux facteurs, notamment culturels (liés à la place de la femme au sein de la société malienne) et économiques (coûts directs et d'opportunité de l'éducation). Ainsi, pour les enfants de 7 à 14 ans, un écart de fréquentation scolaire de 6% entre les filles et les garçons (56,7% contre 50,9%) est observé<sup>3</sup>. L'accès à l'enseignement fondamental et secondaire est également conditionné par le lieu d'habitation. Les enfants en milieu rural ont moins de chance d'aller à l'école que ceux vivant en ville. On note un écart de la fréquentation scolaire de 30,3% entre milieu urbain et rural (75,2% contre 44,9%). Les distances entre lieu d'habitation et école peuvent être en milieu rural de plusieurs dizaines de kilomètres. Pour l'accès aux infrastructures scolaires, de fortes inégalités régionales sont constatées, notamment pour les régions de Kidal, Tombou-

tu et Kayes. La distance moyenne à l'école fondamentale de premier cycle est pour ces régions supérieure à 2 km. Pour le second cycle de l'école fondamentale, la distance moyenne est proche de 10 km pour la région de Kayes. De manière surprenante, la région de Gao présente des distances moyennes à l'école relativement faibles, de l'ordre du km. Pour l'accès aux autres infrastructures (eau, santé, transports...), nous retrouvons ces inégalités régionales. 90% de la population dispose d'une source d'eau potable à moins d'1 km mais la distance peut être bien plus importante dans les régions désertiques. Dans la région de Kidal, la distance à la source d'eau potable est en moyenne de 5,9 km.

Outre la question de la fréquentation scolaire, réside également celle de la réussite scolaire. Même si cette dimension ne fera pas l'objet de notre étude, il est important d'avoir à l'esprit que les systèmes éducatifs des pays en voie de développement présentent de forts taux de redoublements. En l'occurrence, ce taux est de 17% dans l'enseignement primaire au Mali en 2005 (source Unesco). Le taux de redoublement

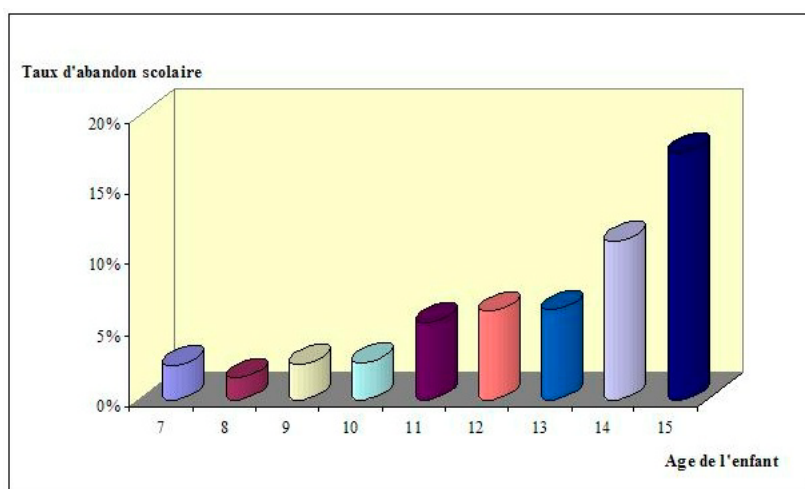
ne peut être calculé directement à partir de l'enquête ELIM. Si les taux de redoublements sont relativement importants, les taux d'abandon scolaire (i.e. le ratio de personnes ayant abandonné l'école après avoir été scolarisées) sont plus faibles : une fois scolarisé, l'enfant a tendance à rester scolarisé jusqu'à 14 ans. Néanmoins, deux ruptures sont constatées, à 11 ans et à 14-15 ans (cf. figure 2).

In fine, malgré une augmentation importante des effectifs durant les dernières décennies et des moyens mis en œuvre afin d'augmenter l'offre d'éducation, le bilan est contrasté. En effet, le ratio d'élèves par classe au primaire (56%) et le taux de redoublement (17%) restent élevés. La qualité de l'enseignement, l'égalité d'accès à l'école pour les filles, ainsi qu'une augmentation des taux de scolarisation restent des défis majeurs et cruciaux pour le Mali.

## 2. Revue de littérature

L'existence d'un lien entre d'une part, l'environnement culturel, entendu comme stock d'éducation,

### F2 Abandon scolaire en fonction de l'âge de l'enfant



Source : ELIM-2006

<sup>3</sup> Les écarts constatés sont significatifs. Les tests d'égalité de proportion classiques supposent que les observations sont indépendantes entre elles. Ce qui revient à ne pas prendre en compte la dimension " ménages " : au sein d'un même ménage, les observations ne peuvent pas être considérées indépendantes. Une solution personnelle pour prendre en compte ce facteur a été de régresser, par les MCO, le fait d'être ou non scolarisé sur la variable d'intérêt dichotomique (le genre, le milieu...) en corrigeant la matrice de variance-covariance par l'auto-corrélation au niveau des ménages. Le coefficient associé à la variable d'intérêt correspond alors à la différence des moyennes et un test de significativité (tenant compte de la correction de la variance) permet de conclure. Dans nos analyses, l'interprétation des résultats des tests reste inchangée.

et économique du ménage et d'autre part, la scolarisation des enfants qui le composent semble intuitive de prime abord. Néanmoins, les mécanismes en œuvre sont complexes, en témoigne l'abondance de la littérature sur le sujet.

Cette littérature s'appuie en premier lieu sur les travaux de Gary Becker sur la théorie du capital humain et le rôle de la famille. Dans ce cadre, le choix d'investissement en capital humain, par les parents ou les enfants, est fonction d'une analyse des coûts et des bénéfices associés.

Nous détaillerons tout d'abord l'impact du capital économique, puis celui du capital éducatif du ménage sur la scolarisation des enfants. La littérature identifie également de nombreux autres facteurs (notamment les caractéristiques démographiques des familles et des enfants, le milieu de résidence, le genre etc.), introduites comme variables de contrôle dans les modèles.

### **Le rôle du capital économique**

Concernant l'impact du capital économique, la scolarisation génère, en plus des coûts directs (uniformes, matériels scolaires,...), des coûts d'opportunités sur le revenu du ménage. Deux types de coûts d'opportunité peuvent être distingués.

D'une part, lorsqu'il va à l'école, l'enfant coûte à ses parents la valeur de l'activité productive qu'il ne réalise pas. D'autre part, éduquer un enfant peut être coûteux en termes de temps pour les parents : pour amener l'enfant à l'école, l'aider à faire ses devoirs, voire le temps consacré dans le suivi scolaire de l'enfant. Becker et al. (1990) notent ainsi que le temps engendre un coût d'opportunité proportionnel au salaire des parents, dans le cadre d'un arbitrage temporel entre éducation et travail. Cependant, il est important de conserver à l'esprit que cette analyse est menée sur des pays développés présentant déjà un fort taux de scolarisation. Il est évident que la transposition de cette théorie dans le cadre de pays en voie de développement tels que

le Mali ne peut se faire directement. Ainsi, le suivi scolaire ne peut être le fait que des parents éduqués eux-mêmes, ce qui restreint la population impactée par ce phénomène. Par ailleurs, ce rôle de suivi du parent peut être suppléé par les enfants aînés scolarisés qui eux-mêmes voient ainsi restreint le temps alloué à leur travail scolaire.

En contrepartie de ces coûts d'opportunité, le capital humain accumulé au travers de l'éducation offre des flux de revenus futurs plus élevés au ménage. Toutefois, même lorsque les gains futurs de l'éducation font plus que compenser ses coûts présents, le faible développement et les imperfections du marché du crédit peuvent obliger certains ménages à renoncer à éduquer leur enfant. Dans un tel contexte économique, certains ménages subissent des contraintes de liquidité : leur incapacité à emprunter de l'argent les contraint à posséder un patrimoine minimal à chaque instant. Les ménages les plus pauvres se retrouvent alors dans l'impossibilité de substituer du revenu futur à du revenu présent, quand bien même ce choix serait optimal. En conséquence, l'enfant n'est pas éduqué. Cette situation peut expliquer également la différence de scolarisation observée entre genres. Odaga et Heneveld (1995) notent ainsi que dans les pays d'Afrique sub-saharienne, les filles sont susceptibles de tomber enceinte ou de se marier avant d'avoir terminé leurs études. Une fois mariées, elles quittent le ménage et les parents ne profitent pas des bénéfices du capital humain accumulé. Les parents vont ainsi privilégier l'éducation des garçons. Au contraire, Wolfe et Behrman (1984) montrent qu'en zone rurale au Nicaragua, les garçons fréquentent moins l'école que les filles. En effet, pour les travaux de nature agricole, les coûts d'opportunité associés à la fréquentation sont plus importants pour les garçons.

Néanmoins, être scolarisé n'est pas une condition suffisante à l'accumulation de capital humain. En particulier, les conditions offertes à l'enfant pour étudier sont primordiales dans sa réussite. Ainsi, les condi-

tions de vie auront un impact positif sur le choix de scolarisation, en augmentant les revenus engendrés par l'éducation et donc l'incitation à scolariser son enfant. Le confort matériel ayant un coût, cet effet lie donc positivement la richesse et la scolarisation de l'enfant.

Les études empiriques montrent sans surprise que le revenu des parents a un impact positif sur la scolarisation des enfants. Ceci est d'autant plus vrai dans le cadre des économies en développement où les contraintes de richesse sont importantes. Ainsi, des études menées au Kenya (Buchman 2000), en Ouganda (Okumu 2008) et au Mexique (Bowman 1984) font toutes des revenus du ménage, l'un des déterminants le plus important dans la décision de scolarisation.

Des politiques en faveur de l'éducation mises en œuvre dans des pays en développement soulignent également l'importance du capital économique. Dans ce contexte, l'étude menée par Duflo, Dupas et Kremer montre que le financement de l'uniforme au Kenya a permis d'améliorer de près de 10% la fréquentation scolaire et de diminuer l'absentéisme d'un tiers.

De même, la plupart des programmes visant à subventionner la participation des enfants au système éducatif engendrent un impact positif sur la scolarisation. Ainsi, la mise en place d'allocations pour les parents envoyant leurs enfants à l'école (programme Progreso au Mexique en 1997) ou la délivrance de bourses au mérite (étude de Miguel, Kremer et Thorntorn au Kenya en 2004) engendrent dans les deux cas une augmentation de la fréquentation scolaire des enfants concernés. Ces résultats confirment ainsi l'existence probable d'une contrainte de liquidité dans de nombreux ménages.

### **Le rôle du capital culturel**

Concernant l'impact du capital éducatif des parents, les études empiriques montrent que l'éducation des parents est un déterminant essentiel de la fréquentation scolaire. Les travaux de Behrman et

al. (1999) valident ainsi l'effet positif de l'éducation des parents sur la scolarisation des enfants. En effet, la capacité d'un enfant à réussir à l'école dépend de ses conditions de vie au sein de la cellule familiale. Nous avons souligné l'importance du confort matériel. La capacité des parents à assister l'enfant dans ses études importe également. Plus généralement, le capital humain de l'enfant peut être analysé comme étant le produit d'un processus de production à facteurs complémentaires avec comme intrants le capital humain des parents d'une part, et les heures passées à l'école d'autre part. En ce sens, le rendement de la scolarisation sera plus élevé pour un enfant naissant dans un foyer riche en capital humain. Les études sociologiques portant sur la mobilité intergénérationnelle dans les pays développés (C.Thélot, 1982) montrent clairement le rôle d'externalités positives jouées par la famille. Ainsi, il est probable que si ses parents ont fait des études, l'enfant développe à leur contact ce goût pour le savoir.

De plus, les choix de scolariser un enfant pour les parents reposent en grande partie sur la perception des rendements espérés du capital humain et ici plus particulièrement de l'éducation. La précision avec laquelle les parents évaluent ce rendement dépend de l'information qu'ils ont accumulée au cours de leur vie, et en particulier au cours de leurs études. Si l'on suppose que la méconnaissance des rendements de l'éducation amène à sous-estimer l'appréciation qu'un agent en a, alors les parents les plus éduqués enverront leurs enfants à l'école car ils ont la capacité d'apprécier à leur juste valeur les gains de l'éducation.

L'impact positif de la connaissance des rendements du système éducatif intéresse depuis peu les programmes d'aide à la scolarisation. Ainsi, il s'avère que les parents et enfants sous-estiment toujours les rendements de l'éducation. Des études de Jensen en République dominicaine (2004), et de N'Guyen à Madagascar (2007) montrent qu'informer les enfants sur les rendements futurs des diplômes, engendre toujours une augmenta-

tion de la fréquentation scolaire et une diminution de l'échec.

De manière plus large, les travaux de Bowman (1984) sur des données indiennes, brésiliennes et mexicaines montrent l'influence positive de la présence d'une radio dans le ménage. Celle-ci permettrait de capter le degré d'ouverture sur le monde extérieur, facilitant la perception des bénéfices de l'école.

Même si la littérature suggère de nombreux liens entre capitaux économique et culturel et probabilité de scolarisation des enfants, les différents effets évoqués vont dans le même sens: un enfant aura d'autant plus de chance d'aller à l'école que sa famille est riche et éduquée. Toutefois, les effets richesse et éducation étant fortement imbriqués dans la théorie, il apparaît à la fois important et intéressant de tenter de démêler ces deux dimensions et ainsi d'en mesurer l'importance relative.

*Cette section nous a permis d'analyser la pertinence de l'étude, au regard du contexte éducatif et social du Mali et des éléments issus de la littérature socio-économique. Notre population d'intérêt est la tranche des enfants de 6 à 15 ans, pour lesquels la scolarisation de base ne concerne encore que la moitié des enfants. Les âges extrêmes seront traités de manière particulière car ils ne font pas partie de l'âge légal de scolarisation obligatoire. Nous allons maintenant définir les indicateurs de mesure du capital économique et culturel à partir des données disponibles.*

## IV. Analyses descriptives

### 1. Descriptif des données

Les données sont issues de l'enquête ELIM 2006 (Enquête Légère Intégrée auprès des Ménages), réalisée par la Direction Nationale de la Statistique et de l'Informatique du Mali. Cette enquête a bénéficié du soutien de la Banque Mondiale et de l'Union Européenne. Elle a pour but d'établir les principaux indicateurs de bien-être des différents groupes socio-économiques de la population, ainsi que les données relatives à la pauvreté monétaire. Elle alimente en statistiques le Document de Stratégie pour la Croissance et la Réduction de la Pauvreté (DSCR). Elle couvre les domaines suivants: la démographie, la santé, l'éducation, l'emploi, les caractéristiques du logement, la pauvreté subjective, la démocratie et la gouvernance.

L'échantillon complet contenait 40810 individus. Notre population d'intérêt, les enfants de 6 à 15 ans, en contient 12412 dont 1448 de 6 ans et 1228 de 15 ans. Les enfants de 11 ans sont moins bien représentés, avec 918 individus. A partir de l'échantillon complet, de nombreuses variables ont été créées, y compris différents indicateurs du capital économique et éducatif. Ces variables, ainsi que les choix de partition en classes, sont décrits plus précisément dans les analyses descriptives et les spécifications des modèles mise en place.

Le tableau 1 résume les caractéristiques de l'échantillon total et de notre population d'intérêt. Les 12412

#### T<sub>1</sub> Caractéristiques de l'échantillon ELIM-2006

Table totale	Nombre de grappes	749
	Nombre de ménages	4494
	Nombre d'individus	40810 (12.32 millions)
Population d'intérêt	Nombre d'enfants de 6 à 15 ans	12412 (3.70 millions)
	Nombre de ménages	3705



individus de 6 à 15 ans représentent 3.70 millions d'habitants (sur une population totale de 12.32 millions).

En annexe 1., les tableaux 13 et 14 décrivent la distribution des principales variables discrètes et continues. Nos principales variables discrètes sont le milieu, le genre de l'enfant, le fait qu'il travaille ou non, et la scolarisation de l'enfant (notre variable d'intérêt). Parmi les variables continues, on trouve les niveaux d'étude au sein du ménage (chef de ménage et conjoints) et le proxy de la richesse du ménage à savoir le montant des dépenses totales et alimentaires. L'analyse de la distribution des variables est particulièrement nécessaire pour la compréhension des analyses descriptives. Ainsi, la répartition entre genre est inégale dans notre échantillon, 46% de filles contre 54% de garçons. Les variables continues présentent des écarts-types élevés, synonyme de variables fortement dispersées.

## 2. Le capital éducatif

Tout d'abord, il est à noter que, dans la suite de l'étude, les expressions capital éducatif et capital culturel renvoient à la même notion, i.e. le niveau d'éducation du ménage.

Pour mesurer le capital éducatif d'un ménage, une première idée est de prendre en compte le nombre d'années d'études validées par ses membres. Nous constatons de nouveau une différence fortement significative entre milieux. Pour les individus de plus de 15 ans, le niveau d'étude moyen est de 0.9 en milieu rural contre 3.8 en milieu urbain.

Cependant, le niveau d'étude ne mesure qu'imparfaitement le capital éducatif. En effet, le fait d'avoir validé les premières années d'études de l'école fondamentale n'assure pas de savoir lire et écrire. Après avoir validé au moins le premier cycle, soit 6 années ou plus d'école, tout le monde sait lire et écrire. La figure 3 nous montre qu'au fur et à mesure des 5 premières années d'étude, le taux d'alphabétisation progresse mais qu'il n'est pas de

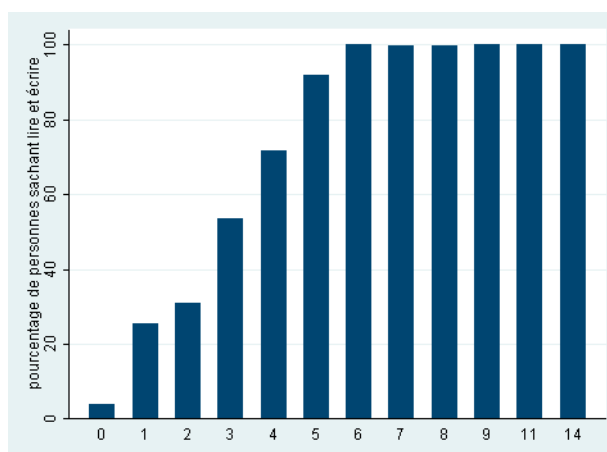
100%. Entre la deuxième année et la cinquième année d'étude, chaque année d'étude supplémentaire fait progresser le taux global d'environ 20%. Ainsi, afin de bien appréhender le capital éducatif, il est nécessaire de tenir compte à la fois du niveau d'étude et de l'alphabétisation des membres du ménage dans nos analyses ultérieures. Après 6 années d'étude validées, ces caractéristiques sont par contre parfaitement corrélées.

Il faut également tenir compte de la répartition du capital éducatif dans le ménage. Quel(s) membre(s) aura une influence spécifique sur la fréquentation scolaire des enfants ? Intuitivement, le capital éducatif du chef de ménage apparaît comme déterminant. Or, le niveau d'étude du conjoint peut également avoir

une influence sur la fréquentation scolaire de l'enfant. Par exemple, les filles pourraient être plus influencées par le capital éducatif de leur mère. D'autre part, un effet d'entraînement de la part d'autres membres du ménage, tels que les frères et sœurs plus âgés, pourrait être déterminant. Analysons cette répartition en regardant le niveau maximal d'étude atteint, pour les ménages de notre population d'intérêt, soit parmi le chef de ménage, son ou ses conjoints (lorsqu'ils ne sont pas décédés), i.e. l'ensemble des membres de plus de 15 ans ou de 18 ans.

À partir du tableau 2, nous constatons que le chef de ménage a globalement un capital éducatif plus important que ses conjoints, mais que le capital éducatif des autres membres du ménage, notamment

### F3 Alphabétisation en fonction du nombre d'années d'étude



Source : ELIM-2006

### T2 Niveau d'étude atteint des membres du ménage (en %)

	Chef de ménage	Conjoints(s)	Membres de plus de 18 ans	Membres de plus de 15 ans
Aucun	79.0	86.0	59.3	49.6
Ecole fondamentale 1 (primaire)	5.8	6.3	10.1	11.6
Ecole fondamentale 2 (collège)	8.7	5.2	17.0	25.0
Lycée et plus	6.5	2.6	13.7	13.8
Total	100	100	100	100

Source : ELIM-2006

jeunes, apparaît supérieur. Pour les ménages dont aucun membre n'a fréquenté l'école, un écart de 10 points est mis en évidence selon que l'on prend en compte uniquement les membres de 18 ans et plus, ou ceux de plus de 15 ans. Cette situation peut être due aux récentes politiques de scolarisation au Mali mais également au fait que certains jeunes partent vraisemblablement du foyer familial à la fin du second cycle de l'enseignement fondamental (collège) soit pour poursuivre leurs études, soit pour rechercher du travail. Pour mesurer un éventuel effet d'entraînement des frères et sœurs, il apparaît donc plus pertinent de prendre en compte l'ensemble des membres de plus de 15 ans et de considérer la variable mesurant le maximum du niveau d'étude atteint au sein du ménage plutôt que celui du seul chef de ménage.

### 3. Le capital économique

L'analyse de la distribution de la consommation des ménages (cf. figure 4, pour les enfants de 7 à 14 ans) fait apparaître une relative concentration de celle-ci autour d'un niveau modal relativement faible. Cette concentration vers le bas des niveaux de consommation concorde avec le caractère généralisé de la pauvreté mis en évidence dans la section contextuelle précédente.

Pour expliquer les différences de capital économique entre les ménages, nous avons analysé la relation entre le niveau d'études du chef de ménage et les dépenses de consommation. Pour ce faire, le niveau de consommation médian pour chaque niveau d'étude a été étudié (cf. figure 5). Le choix de la médiane, plutôt que la moyenne, permet de contrôler l'analyse par rapport aux valeurs extrêmes de la consommation des ménages.

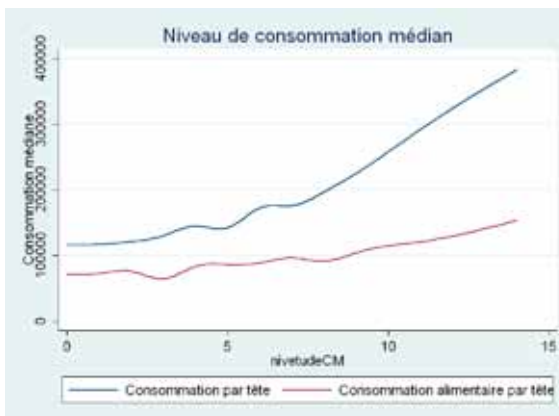
L'impact de la scolarité des parents sur la consommation du ménage apparaît seulement après un certain niveau d'études. Ainsi, celui-ci décolle après 6 années d'étude et atteint un niveau particulièrement élevé pour les plus diplômés. Deux éléments peuvent expliquer ceci.

## F4 Distribution de la consommation totale par tête (en FCFA par an)



Source : ELIM-2006

## F5 Niveau de consommation médian (en FCFA par an) en fonction du niveau d'étude



Source : ELIM-2006

Tout d'abord, une première explication peut venir du fait que les premières années de scolarité n'apportent qu'un faible gain en terme de capital humain : ceci confirme le fait mis en avant qu'il faudrait au moins 6 années d'étude pour savoir correctement lire et écrire.

Ensuite, l'aspect quasi-exponentiel de la consommation peut provenir de caractéristiques propres aux données. En effet, celles-ci sont censurées vers le haut dans le sens où le niveau maximum correspond aux individus ayant fait des études supérieures : celles-ci ne sont donc pas seulement une année d'étude supplémentaire mais sans doute plus. De même, l'avant-dernière catégorie correspond au niveau

d'étude "secondaire" et possède donc les mêmes écueils concernant la surévaluation de cette "année" d'étude supplémentaire.

En France, la consommation alimentaire représente 13,4% de la consommation totale des ménages en 2007 (INSEE). Elle est en moyenne de 57% au Mali. D'un point de vue micro-économique, la caractéristique de saturation de la consommation est importante pour les biens alimentaires. La pente de la consommation alimentaire est ainsi plus faible que celle de la consommation totale. Ainsi, le ratio des dépenses d'alimentation sur les dépenses totales pourrait également représenter une mesure du capital économique du ménage.

#### 4. La fréquentation scolaire selon nos deux dimensions

Dans un premier temps, nous observons, sur les figures 6 et 7, la forme de la relation entre nos deux variables d'intérêts et notre variable à expliquer. Pour cela, on relie la probabilité pour un enfant d'être scolarisé suivant la consommation par tête du ménage et le niveau d'étude maximal atteint par un adulte. La variable de consommation étant continue, on a recours à une régression non-paramétrique<sup>4</sup>. On note que la forme de la relation semble concave dans les deux cas.

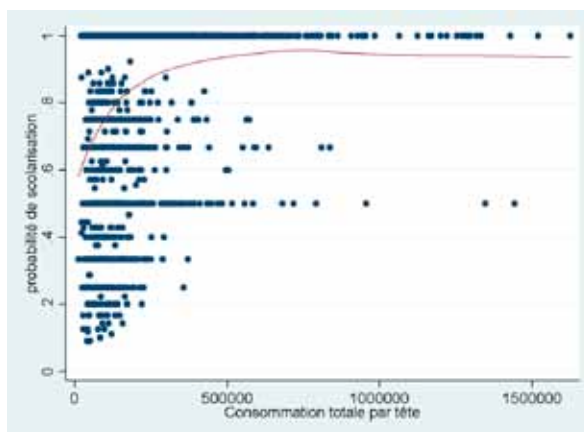
Ainsi, l'effet marginal de nos deux capitaux culturel et économique semble fortement diminuer avec leur niveau, ce qui est cohérent avec l'hypothèse d'existence de fortes contraintes de liquidité.

Pour approfondir cette analyse, on s'intéresse maintenant aux différentes variables caractérisant ces capitaux, en présentant leur effet marginal sur la scolarisation dans les figures 8 et 9. Pour analyser la relation entre la fréquentation de l'école et nos variables, il est judicieux de calculer les taux de fréquentation scolaire par modalités et de les comparer à la moyenne générale.

Pour la mesure du capital économique, on trouve ainsi les relations attendues quel que soit l'indicateur retenu : la consommation totale par tête, la consommation alimentaire par tête, la part de la consommation totale ou l'indicateur de pauvreté absolue, calculé par la DNSI. Pour le capital éducatif, l'effet des études supérieures sera plus marqué pour le chef de ménage que pour l'ensemble des membres de plus de 15 ans. Ce sera l'inverse pour l'effet de l'absence d'études. On note également un effet différencié du primaire et du collège en prenant en compte l'ensemble des membres du ménage. Dans les figures 8 et 9 apparaissent en rouge le taux de fréquentation scolaire moyen (pour les enfants de 7 à 14 ans). Les points représentent

F6

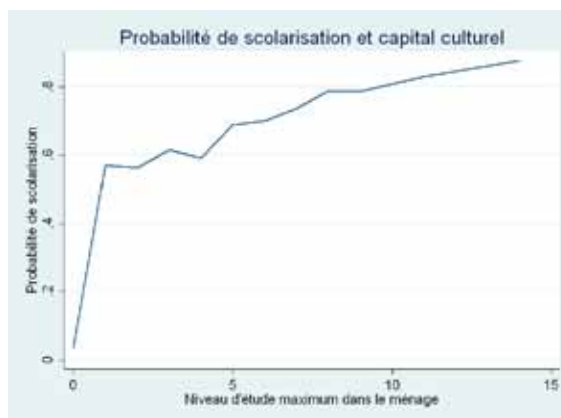
#### Consommation totale par tête et probabilité de scolarisation



Source : ELIM-2006

F7

#### Niveau maximal d'étude et probabilité de scolarisation



Source : ELIM-2006

le taux de fréquentation scolaire par modalités. Pour les deux premières variables, les modalités représentent les quartiles de la distribution. Les modalités *Int faible* (position intermédiaire faible) et *Int fort* (position intermédiaire forte) représentent ainsi le deuxième et troisième quartile. Plus le capital économique est important, plus la fréquentation scolaire observée est forte.

Ces analyses ont été conduites pour l'ensemble des variables de notre étude et sont disponibles en annexe 2. Elles permettent de vérifier que chaque variable a le rôle attendu. Elles mettent également en évidence l'importance des regroupements des

modalités effectués. Par exemple, pour caractériser l'absence du chef de ménage, deux situations sont distinguées selon que l'absence soit supérieure ou non à 6 mois. Si on regroupe ces deux modalités, le taux de fréquentation scolaire est alors supérieur à la moyenne en cas d'absence du chef de ménage, ce qui n'est en réalité le cas que pour une absence inférieure à 6 mois.

Les résultats sont cohérents avec la théorie ou l'intuition pour la majorité des variables. On constate néanmoins quelques résultats surprenants. Ainsi, si les orphelins présentent un taux de fréquentation scolaire inférieur à la moyenne,

<sup>4</sup> Dans ce cas, il s'agit d'une régression partitionnée par les moindres carrés ordinaires. Le principe est de multiplier les régressions de la probabilité de scolarisation sur la consommation en découpant notre intervalle. On aboutit ainsi, en agrégeant ces régressions, à une relation non-linéaire entre les deux variables.

les enfants ayant perdu un de leurs deux parents seulement fréquentent plus souvent l'école. L'effet de la taille du ménage et du nombre d'enfants n'est également pas clair. Mais le nombre d'enfants au sein du ménage est une variable dont l'effet théorique attendu est complexe.

**Conclusion partielle :** ces analyses nous permettent de sélectionner les indicateurs qui seront pris en compte pour le capital économique et éducatif. Le capital économique sera mesuré par la dépense totale du ménage par tête car cette variable est la plus discriminante. Pour le capital éducatif, nous tiendrons compte de deux variables, le niveau d'étude maximal du ménage (pour les individus de plus de 15 ans) et la capacité à lire et à écrire du chef de ménage.

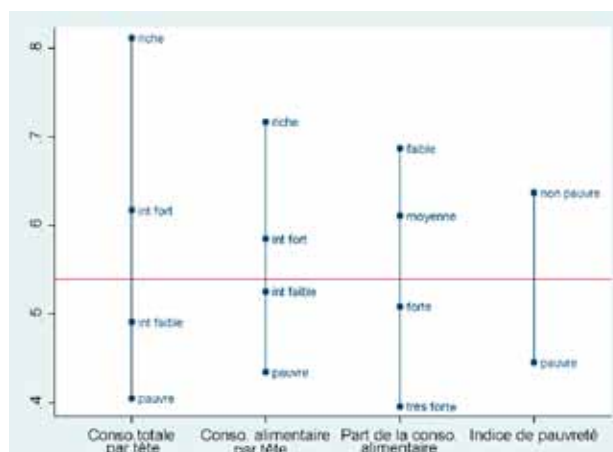
## V. Modelisation économétrique

Dans cette section, nous étudions dans quelle mesure les deux dimensions de notre analyse, à savoir le capital économique et éducatif du ménage, expliquent la scolarisation des enfants au Mali. Cependant, d'autres facteurs affectent également la décision de scolarisation de l'enfant. Nous nous emploierons à identifier ces facteurs et à les utiliser comme variables de contrôle dans notre analyse. Pour ce faire, nous disposons, de par l'enquête ELIM, de multiples variables qui nous permettent de contrôler l'aspect offre d'éducation, les inégalités en termes de situations géographiques, d'accessibilité aux infrastructures et d'autres contraintes extérieures pesant sur les ménages dans leur choix de scolariser leurs enfants.

### 1. Cadre théorique

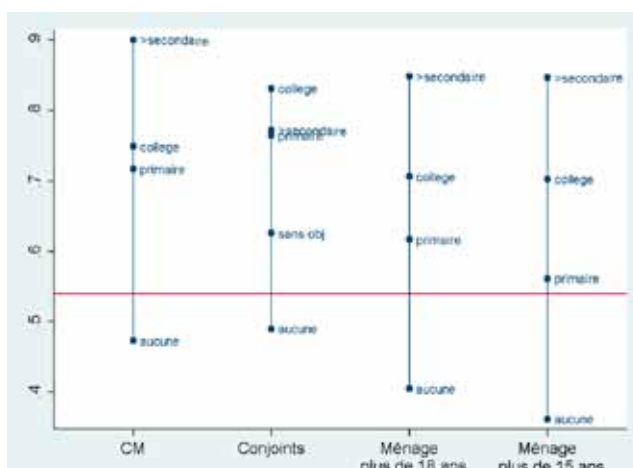
Notre variable d'intérêt étant de nature dichotomique, un modèle non linéaire est ici particulièrement adapté (logit ou probit). Notre variable dépendante, notée  $EC_{ij}$ , prend pour valeur 1 lorsque l'enfant  $i$  du ménage  $j$  fréquente l'école, 0 sinon.

## F8 Fréquentation scolaire et capital économique



Source : ELIM-2006

## F9 Fréquentation scolaire et capital éducatif



Source : ELIM-2006

En notant  $m$  le nombre de ménages et  $n$  le nombre d'enfants au sein de notre enquête, un enfant est identifié par son numéro  $i \in [1, n]$  et par son ménage d'appartenance  $j \in [1, m]$ .

$$\mathbb{P}(EC_{ij}) = F(\text{eco}_j, \text{cul}_j, \text{Cenf}_{ij}, \text{Cmen}_j, \text{Ccom}_j), \forall (i, j) \in [1, n] \times [1, m] \quad (1)$$

où les **variables d'intérêt** sont :

- $\text{eco}_j$ , les caractéristiques du capital économique du ménage (niveau de consommation totale par tête et, dans une moindre mesure, la CSP du chef de ménage);
- $\text{cul}_j$ , les caractéristiques du capital éducatif du ménage (niveau d'éducation maximum au sein du

ménage pour les adultes de plus de 15 ans et, capacité de lecture et d'écriture du chef de ménage)

et les **variables de contrôle** :

- $\text{Cenf}_{ij}$ , caractéristiques propres à l'enfant (genre, âge, absence du ménage au sein de l'année écoulée, travail au cours de l'année écoulée, lien de parenté avec le chef de ménage, parents vivants ou décédés);
- $\text{Cmen}_j$ , les caractéristiques propres au ménage (âge et genre du chef de ménage, absence du chef de ménage au cours de l'année écoulée, nombre d'enfants de moins de 15 ans, nombre de conjoints);



- Ccom<sub>j</sub>, les caractéristiques propres à la communauté d'appartenance (milieu, région, distance à l'eau potable, aux infrastructures de transports et scolaires).

La population d'intérêt est celle des enfants de 6 à 15 ans. Pour les âges extrêmes (6 et 15 ans), une indicatrice d'appartenance est définie afin de tenir compte de la spécificité de la fréquentation scolaire à ces âges.

Les parties suivantes détaillent la construction du modèle. Un modèle de base est d'abord mis en place, mais dont l'ajustement n'est pas bon. Le modèle est ensuite amélioré, pour tenir compte du plan de sondage (corrélation au sein des ménages), des variables endogènes et des données aberrantes.

## 2. Pondérer ou non le modèle?

Les données employées sont issues d'une enquête réalisée par sondage<sup>5</sup>. Les analyses descriptives précédentes ont pris en compte les poids d'enquête. Pour l'étape de modélisation, la question est plus délicate tant d'un point de vue pratique que théorique. En effet, si le modèle est pondéré, les écart-types calculés par les logiciels en tenant compte des poids sont en général non convergents. Des méthodes de retraitement (bootstrap notamment) sont alors nécessaires<sup>6</sup>. D'un point de vue théorique, la connaissance du plan de sondage et des méthodes de traitement de la non-réponse sont nécessaires pour l'analyse. D'après les documents d'enquête, le plan de sondage s'est appuyé sur le milieu, la taille des ménages et le budget des familles, variables de contrôle dans notre modèle. De plus, il n'est pas fait mention de méthodes de pondération ex-post, ni de traitements de la non-réponse. Dans ce cas, l'estimation du modèle logistique dépendant uniquement de la

loi conditionnelle  $Y$  sachant  $X$  (avec  $Y$  la variable dépendante et  $X$ , les variables explicatives), l'estimateur non pondéré sera plus précis. Les deux estimateurs, pondérés et non pondérés, seront convergents. Cette analyse nous conduit à retenir l'estimateur non pondéré.

## 3. Quelle spécification pour les variables d'intérêt ?

Pour les variables continues d'intérêt, la partition en classes a été établie soit à partir des quartiles (capital économique), soit à partir du contexte éducatif du Mali (capital éducatif). Pour les variables de contrôle, nous avons laissé les variables en continu (ce qui revient à enlever les valeurs manquantes pour les variables de distance). Pour les variables qualitatives, une analyse économique des taux de scolarisation par classes a été effectuée pour valider les regroupements.

L'estimation de ce modèle de base fournit les paramètres suivants, pour les principales variables<sup>7</sup> :

Pour approfondir la modélisation des effets des variables principales, en particulier la recherche des effets d'interactions, nous avons également estimé des modèles avec les variables en continu. Le découpage par classes peut également s'avérer arbitraire, notamment pour la variable de consommation. En effet, bien que le découpage du capital culturel (dans cette partie, mesuré par le niveau d'étude du chef de famille) suit l'échelonnement du système scolaire, celui de la consommation repose sur une simple division de notre échantillon en quartiles. Ainsi, on peut s'interroger sur la pertinence de ce choix et s'il ne perturbe pas les analyses que l'on peut tirer de nos estimations.

Aussi, nous testons ici des spécifications alternatives, notamment en

utilisant la continuité naturelle de nos deux variables d'intérêt et en intégrant une variable croisée entre le capital culturel et économique. L'idée est ici de capter un effet généré par une première expérience sur le marché du travail du chef de ménage. En effet, on peut supposer qu'un adulte ayant réussi financièrement sans faire d'étude, ou ayant échoué sur le marché du travail malgré un niveau d'étude élevé, aura une opinion pessimiste des rendements espérés de l'éducation. Il aura dans ce cas tendance à moins inciter ses enfants à aller à l'école, ce qui se traduira par un signe négatif.

Nous cherchons à tester la possible existence d'un lien non-linéaire de nos deux variables d'intérêt. Concernant le niveau d'éducation, celui-ci étant limitée à 10 modalités, nous utilisons un polynôme d'ordre deux, en supposant qu'il permet une description suffisante d'une relation qui serait concave ou convexe. Pour le niveau de consommation, il semble possible que la forme du lien varie selon le profil de consommation. Ainsi, on utilise, dans un premier temps, un polynôme d'ordre 3. Deux modèles sont ainsi produits.

Une première estimation avec cette spécification (modèle (1)) montre que l'effet de retournement de la consommation ne se produit que pour des niveaux de consommation très élevés, mettant en évidence la relative inutilité d'un polynôme d'ordre 3. Ainsi, nous effectuons, dans un second temps, une estimation avec deux polynômes d'ordre 2 pour chacune des variables d'intérêts (modèle (2)). Le tableau 3 reporte les coefficients estimés par STATA.

Deux remarques importantes peuvent être faites. Tout d'abord, la variable croisée n'est significative que dans le cas où l'on retire l'ordre 3 du polynôme. On note dans ce cas que le degré 2 est non-significatif,

<sup>5</sup> Cette partie s'appuie sur une note de X. D'Haultfoeuille et L. Davezies. Un test d'Hausman a été mis en place pour tester la compatibilité du processus de sélection et du modèle paramétrique retenu. L'information sur les retraitements d'enquête étant imprécise, les résultats restent néanmoins d'une interprétation délicate et ne sont pas joints.

<sup>6</sup> STATA propose néanmoins des options de prise en compte de plan de sondage complexe (svy). Différentes méthodes de corrections de variance sont alors proposées. Certains tests ne peuvent par contre plus être implémentés (gof test).

<sup>7</sup> Pour les variables dichotomiques (LireCM, Absence, Travail), la modalité 1 correspond à Oui. Le capital éducatif est divisé en 4 classes : Pas d'études, Ecole fondamentale du 1er cycle, Ecole fondamentale du second cycle, Lycée et études supérieures. Le capital économique est divisé en 4 classes (quartiles de la distribution), qui sont nommées pour simplifier la lecture des tableaux : Très pauvre, Pauvre, Riche, Très Riche. La catégorie socio-professionnelle est divisée en 5 classes : Cadre (catégorie supérieure), Employé, Profession libérale, Catégorie inférieure, Sans emploi.

traduisant une spécification linéaire de la consommation. Ainsi, le poids important des niveaux de consommation élevés perturbe l'estimation de l'effet pour les ménages plus représentatifs de l'échantillon. En comparaison avec la spécification du modèle (1), la seconde semble donc bien moins adaptée pour caractériser cette relation. En conclusion, il ne semble pas que cette variable croisée ait un rôle spécifique dans la fonction de demande en éducation des ménages. Nous avons donc maintenu notre partition en classes des variables d'intérêt.

#### 4. Résultats du modèle simple

Le test de nullité globale des coefficients est rejeté. Le pseudo-R2 est quant à lui relativement faible (0.26) ce qui est souvent le cas des modèles micro-économétriques. Plus intéressante est l'étude de la significativité des paramètres. Les coefficients associés au niveau d'éducation et aux dépenses de consommation totale par tête sont croissants et tous significatifs. Par contre, le fait de savoir lire et écrire (pour le chef de ménage), caractéristique secondaire du capital éducatif, n'a pas d'influence contrairement à ce que pouvait laisser penser les analyses descriptives. Pour la CSP, caractéristique secondaire du capital économique, on constate que les coefficients sont décroissants par rapport à la situation de référence (sauf pour les deux dernières classes), à savoir la catégorie supérieure. Néanmoins, l'interprétation est délicate notamment pour la troisième catégorie (*profession libérale* ou *pour compte propre*) qui est majoritaire dans la base de données.

Il est également intéressant d'analyser si la différence entre les coefficients est significative, pour le niveau d'éducation et les dépenses totales. Des tests de Wald<sup>8</sup> ont été mis en place. Tous les tests présentent des p-value inférieurs à 0.001. Les différences sont donc significatives. Pour les autres variables de contrôle, le genre a une influence

### T<sub>3</sub> Modèle initial

	(1) Non pondéré	(2) Pondéré
Pas d'études	Ref	
Fondamentale 1	0.843*** (12.15)	0.884*** (10.82)
Fondamentale 2	1.398*** (23.10)	1.370*** (18.43)
Lycée et plus	1.764*** (19.54)	1.650*** (14.18)
lireCM	-0.013 (-0.21)	-0.009 (-0.11)
Très pauvre	Ref	
Pauvre	0.155*** (2.76)	0.169** (2.55)
Riche	0.325*** (5.01)	0.405*** (5.26)
Très riche	0.657*** (7.05)	0.659*** (5.37)
Cadre	Ref	
Employé	-0.0481 (-0.34)	0.0455 (0.28)
Libéral	-0.361** (-3.00)	-0.334* (-2.32)
Cat. Inférieure	-0.688*** (-3.72)	-0.481* (-2.18)
Sans Emploi	-0.579*** (-4.35)	-0.485** (-3.07)
Sexe (femme)	-0.399*** (-8.95)	-0.384*** (-7.07)
Age 6	-2.245*** (-24.14)	-2.327*** (-20.40)
Age 15	-0.677*** (-7.42)	-0.723*** (-6.56)
Age	-0.00837 (-0.78)	-0.00233 (-0.17)
Absence	0.0164 (0.19)	-0.241* (-2.18)
NombreEnfants	-0.0143 (-1.94)	-0.0166* (-1.97)
NombreConj	-0.0192 (-0.51)	0.0194 (0.44)
Travail	-1.273*** (-24.54)	-1.269*** (-19.77)
N	12177	12177

Note de lecture : Statistiques de Student entre parenthèses  
Niveaux de significativité : \* : 10% \*\* : 5% \*\*\* : 1%

<sup>8</sup> Le test de Wald s'appuie sur les estimations du maximum de vraisemblance des modèles contraints et non contraints. Sous l'hypothèse nulle où la contrainte est respectée (nullité d'un coefficient, égalité de deux coefficients), la statistique de Wald suit une loi  $\chi^2(p)$  avec  $p$  le nombre de contraintes.

notable. Alors que l'âge n'est pas significatif, les indicatrices pour 6 ans et 15 ans le sont fortement.

Pour tester la qualité de l'ajustement du modèle, nous mettons en place le test "goodness-of-fit" de Hosmer et Lemeshow. Ce test consiste à vérifier que les résultats prédits par le modèle sont cohérents avec les résultats observés. Pour cela, un nombre de classes est défini à partir de la distribution de la probabilité prédite d'être scolarisé  $P(Y = 1|X)$  (le même nombre d'observations est présent dans chaque classe). Sous l'hypothèse nulle que le modèle est bien ajusté, la statistique

$$\sum_{classes} \frac{(O-E)^2}{E}, \text{ avec } O \text{ le résultat}$$

observé et E le résultat prédit, sera proche de 0 et suivra une loi du  $\chi^2$  à (Nombre de classes - 2) degrés de liberté. Quelque soit le nombre de classes choisies (supérieures à 10), l'hypothèse nulle d'un bon ajustement est fortement rejetée.

Quelles peuvent être les causes de ce mauvais ajustement ? C'est ce que nous allons maintenant étudier.

## 5. Tenir compte du plan de sondage pour corriger l'hétéroscédasticité

Les estimations effectuées jusqu'alors reposent sur l'hypothèse d'homoscédasticité des résidus. L'homoscédasticité signifie à la fois que les termes diagonaux de la matrice de variance des résidus sont égaux et que les termes en dehors de la diagonale sont nuls. Cette dernière hypothèse a peu de chance d'être vérifiée puisqu'au sein d'un ménage, les enfants partagent certainement une culture familiale ou tout autre caractéristique non observable et expliquant la scolarisation. La matrice de variance n'est alors plus que diagonale par blocs.

Afin de corriger le biais sur l'estimation de la matrice de variance des résidus nous implémentons la commande *cluster* de STATA (cf. tableau 5).

## T4 Modèle avec des variables d'intérêt continues

	(1)	(2)
EtudeCM	0.234*** (7.82)	0.246*** (8.24)
EtudeCM <sup>2</sup>	-9.45e-3*** (-3.70)	-8.18e-3** (-3.22)
LireCM	-0.182* (-2.01)	-0.181* (-1.99)
ConsoTête	7.08e-6*** (8.64)	3.34e-6*** (9.68)
ConsoTête <sup>2</sup>	-1.09e-11*** (-5.00)	-2.71e-13 (-1.50)
ConsoTête <sup>3</sup>	5.24e-18*** (4.01)	
VariableCroisée	-4.93e-08 (-0.98)	1.73e-7*** (-3.93)
Observations	12177	12177

Note de lecture : Statistiques de Student entre parenthèses  
Niveaux de significativité : \* : 10% \*\* : 5% \*\*\* : 1%

## T5 Régression avec et sans prise en compte de l'autocorrélation au niveau des ménages

EcoleEC	(1) Cluster	(2) Simple
Pas d'études	Ref	
Fondamentale 1	0.858*** (8.68)	0.858*** (12.34)
Fondamentale 2	1.443*** (15.64)	1.443*** (23.32)
Lycée et plus	1.837*** (13.72)	1.837*** (19.78)
LireCM	-0.010 (-0.11)	-0.010 (-0.16)
Très pauvre	Ref	
Pauvre	0.146* (1.75)	0.146*** (2.59)
Riche	0.320*** (3.36)	0.320*** (4.94)
Très Riche	0.700*** (5.45)	0.700*** (7.43)
Observations	12177	12177
Pseudo R <sup>2</sup>	0.265	0.265

Note de lecture : Statistiques de Student entre parenthèses  
Niveaux de significativité : \* : 10% \*\* : 5% \*\*\* : 1%

L'utilisation de cet estimateur abaisse la significativité de l'ensemble des coefficients. Toutefois, contrôler uniquement les conséquences de l'autocorrélation sur l'estimation de la variance est conceptuellement déroutant dans la mesure où dans les modèles non-linéaires, au contraire des modèles linéaires, l'hétéroscédasticité biaise également l'estimation des coefficients. Cela provient du fait que lorsqu'il y a hétéroscédasticité, et en particulier autocorrélation, la vraisemblance est mal spécifiée. Les observations n'étant pas indépendantes, la vraisemblance ne s'écrit pas comme le produit des densités.

Afin de corriger cette erreur de spécification, plusieurs méthodes de traitement de l'hétérogénéité inobservée dans les modèles logit existent. Pour décider laquelle mettre en oeuvre, la première question qui se pose est de savoir si les effets ménages sont corrélés ou non avec les variables explicatives. Supposer l'absence de corrélation étant une hypothèse forte, il est préférable d'estimer un modèle avec effet fixe qui autorise cette corrélation.

Les modèles logit avec effets fixes sont appelés logit conditionnel car la vraisemblance est conditionnée de façon à ne pas dépendre des effets fixes qui ne sont par conséquent pas estimés. Le problème est que, si nous estimons un modèle avec des effets fixes ménage, les coefficients associés aux variables constantes entre les enfants d'un même ménage ne pourront pas non plus être estimés. Or nos variables d'intérêt sont précisément des variables de ménage. Une autre solution (permettant de résoudre ce problème) est d'estimer un logit avec effet aléatoire (cf. tableau 6). Les détails théoriques justifiant ces choix sont présents en annexe 3.

Le modèle se réécrit alors :

$$y_{j,i}^* = x'_{i,j}\beta_1 + z'_j\beta_2 + v_j + e_{i,j} \quad (2)$$

Avec  $y_{j,i}^*$  la variable latente associée à la variable scolarisation de l'enfant  $i$  du ménage  $j$ ,  $z_j$  le vecteur de variables ménages commun à

tous les enfants du ménage,  $x_{i,j}$  un vecteur de variables explicatives propres à chaque enfant,  $v_j$  un effet variable inobservable commun à l'ensemble des enfants du ménage  $j$  et non corrélé avec  $(x_{i,j}, z_j)$  et enfin  $e_{i,j}$  le "vrai" résidu du modèle au sens où il respecte l'hypothèse d'homoscédasticité (cf. tableau 6).

Le signe du coefficient de l'indicatrice d'alphabetisation du chef de ménage est modifié dans le bon sens par l'estimateur avec effet aléatoire. Il reste néanmoins fortement non significatif. D'autre part, les coefficients associés au capital économique et éducatif augmentent.

L'interprétation de ces résultats reste délicate. Ils montrent néanmoins que ces corrections induisent des modifications non négligeables des coefficients. La prise en compte de l'hétéroscédasticité pour des modèles non linéaires reste néanmoins peut abordée dans la littérature économique. Faute de pouvoir trancher de manière claire entre ces différentes solutions, nous avons retenu de corriger l'autocorrélation au niveau des ménages seulement sur le modèle de base.

## 6. L'endogénéité des variables

Nous envisageons deux variables endogènes potentielles : le nombre d'enfants dans le ménage et l'indicatrice travail. Notre suspicion quant au nombre d'enfants est motivée par les résultats des modèles théoriques avec fécondité endogène. Dans le modèle de Becker (1991), les parents choisissent leur nombre d'enfants et le niveau de scolarité de chacun en fonction principalement de leur salaire et des rendements de l'éducation, les deux variables apparaissant comme des substituts stratégiques<sup>9</sup>. Formellement, en supposant que les stratégies d'éducation sont invariantes à l'intérieur d'un ménage, le vrai modèle à estimer a donc la forme suivante :

$$\begin{cases} y_i^* = x'_i\beta + \alpha ne_i + e_i \\ ne_i = \gamma y_i^* + z'_i\mu + v_i \end{cases}$$

Avec  $x'_i$  le vecteur d'observation des variables exogènes de l'enfant moyen du ménage  $i$  plus la constante,  $ne_i$  le nombre d'enfants dans le ménage  $i$ .  $z'_i$  est un vecteur d'observation de variables indépendantes de  $e_i$  et  $v_i$ , dont la constante.

### T6 Régression effets aléatoires versus effets simples

EcoleEC	(1) Effets Aléatoires	(2) Simple
Pas d'études	Ref	
Fondamentale 1	1.392*** (9.78)	0.858*** (12.34)
Fondamentale 2	2.176*** (16.95)	1.443*** (23.32)
Lycée et plus	2.686*** (14.50)	1.837*** (19.78)
LireCM	0.006 (0.05)	-0.010 (-0.16)
Très pauvre	Ref	
Pauvre	0.192* (1.68)	0.146*** (2.59)
Riche	0.446*** (3.46)	0.320*** (4.94)
Très Riche	1.055*** (5.98)	0.700*** (7.43)
Observations	12177	12177
Pseudo R <sup>2</sup>		0.265

Note de lecture : Statistiques de Student entre parenthèses  
Niveaux de significativité : \* : 10% \*\* : 5% \*\*\* : 1%

<sup>9</sup> Deux types de stratégies émergent. Les parents ayant un salaire élevé vont faire peu d'enfants mais hautement éduqués tandis que les parents au salaire faible vont faire beaucoup d'enfants peu scolarisés



Les résidus  $e_i$  et  $v_i$  sont supposés indépendants. Finalement,  $ne_i$  est fonction de  $y_i^*$ , lui-même fonction de  $e_i$ .  $ne_i$  est endogène et omettre cet élément amène à mal spécifier la vraisemblance du modèle ce qui biaise l'estimation des coefficients.

La variable indicatrice travail est également fortement susceptible d'être endogène. En effet, les ressources temporelles de l'enfant sont limitées. A temps de loisir donné, les décisions de scolarisation et de travail sont donc parfaitement liées par cette contrainte. Encore une fois un biais de simultanéité risque d'apparaître. La résolution de ce problème se fait en trois étapes:

### Identification d'instruments

Pour être un instrument, une variable doit présenter la triple propriété d'être extérieure au modèle, d'être corrélée avec la variable à instrumenter sans être corrélée avec le résidu  $e_i$ . Afin d'instrumenter le nombre d'enfants, la variable "nombre de conjoints" semble réunir toutes les conditions. Tout d'abord elle n'est pas significative dans le modèle de base, nous pouvons donc raisonnablement l'en extraire. Ensuite elle est bien corrélée avec le nombre d'enfants. Enfin, même si le choix du nombre de conjoints, comme le choix de scolarisation des enfants, est une décision économique, nous pensons que la première est chronologiquement antérieure à la deuxième. La variable nombre de conjoints est donc a priori indépendante de  $e_i$  dont la réalisation est contemporaine du choix de scolarisation. Pour ce qui est de la variable travail, l'âge de l'enfant présente les qualités d'un bon instrument: l'âge n'est pas significatif dans le modèle de base, il peut être retiré et considéré de fait comme respectant la condition d'exogénéité<sup>10</sup>. D'autre part, l'âge explique le travail. Les deux variables sont fortement et positivement corrélées: assez intuitivement, on observe que la probabilité de travailler augmente avec l'âge.

### Régressions instrumentales

Nous réalisons une régression instrumentale pour chaque variable endogène. Nous régressons le nombre d'enfants du ménage sur l'instrument choisi et sur les autres variables du modèle<sup>11</sup> afin de décomposer la variable nombre d'enfants en une partie exogène par rapport à  $e_i$  et un résidu.

De la même manière, la variable travail est régressée sur l'âge et les autres variables du modèle. La variable travail étant dichotomique, nous estimons la régression instrumentale à l'aide d'un modèle logit. Les variables nombre de conjoints et âge sont significatives respectivement dans chacune des régressions. Cela confirme leur statut de bon instrument.

### Régression augmentée

La dernière étape consiste à estimer le modèle augmenté du résidu des régressions instrumentales. L'estimateur du coefficient estimé associé à  $ne_j$  est désormais convergent. Néanmoins, ce n'est pas le cas pour la variable travail. En effet, bien que cette méthode en deux étapes permette de tester l'endogénéité de la variable, la correction par l'introduction d'un instrument orthogonal aux résidus ne permet pas d'aboutir à des estimations convergentes<sup>12</sup>.

On observe dans le tableau 7 que le résidu de la régression instrumentale du nombre d'enfants n'est pas significatif. Cela signifie que la correction apportée par l'instrumentation n'est pas significative. Cela revient

## T7 Régression augmentée de la scolarisation

EcoleEC	(1) Augmentée	(2) Simple
Travail	-0.854*** (-6.72)	-1.280*** (-25.03)
TravailResid	-0.184*** (-3.64)	
NombreEnfantsTot	-0.022 (-1.54)	-0.016** (-2.40)
NombreEnfResid	0.005 (0.32)	
Pas d'études	Ref	
Fondamentale 1	0.858*** (12.34)	0.842*** (12.13)
Fondamentale 2	1.448*** (22.86)	1.395*** (23.10)
Lycée et plus	1.841*** (19.69)	1.761*** (19.53)
LireCM	-0.011 (-0.18)	-0.015 (-0.24)
Très pauvre	Ref	
Pauvre	0.142** (2.48)	0.154*** (2.74)
Riche	0.316*** (4.81)	0.324*** (5.00)
Très Riche	0.692*** (7.11)	0.657*** (7.04)
Observations	12177	12177
Pseudo R <sup>2</sup>	0.265	0.264

Note de lecture : Statistiques de Student entre parenthèses  
Niveaux de significativité : \* : 10% \*\* : 5% \*\*\* : 1%

<sup>10</sup> Nous avons conscience que l'argumentaire sur l'exogénéité de l'âge par rapport au résidu de la scolarisation est discutable puisqu'il se base sur la non-significativité du coefficient associé à l'âge. Or ce coefficient est a priori biaisé puisqu'il est estimé en présence d'endogénéité.

<sup>11</sup> Dans la régression instrumentale, la variable expliquée a une dimension ménage. Pour des raisons évidentes, elle ne peut donc être expliquée par des variables individuelles telles que la scolarisation puisqu'il y a plusieurs observations par ménage.

<sup>12</sup> Ce problème provient du caractère dichotomique de la variable travail au même titre que la variable de scolarisation. Ainsi, la procédure en deux étapes aboutit à deux normalisations de la variance dans les deux étapes successives. Une procédure régulière pour corriger ce biais dans cette situation consisterait dans une estimation avec un probit bivarié.

à rejeter l'hypothèse d'endogénéité du nombre d'enfants dans le modèle de base. Le coefficient associé au résidu de la régression instrumentale du travail est au contraire significatif au seuil de 1%, on rejette donc à ce seuil l'hypothèse d'exogénéité du travail.

Même si la correction apportée à cette endogénéité de la variable travail n'est pas convergente, on conservera par la suite cette spécification instrumentée, conscients cependant qu'une mauvaise correction n'est pas forcément préférable au problème initial. En ce qui concerne les autres coefficients, les modifications apportées par l'instrumentation sont faibles. On peut noter à propos de la consommation totale que les coefficients des classes 2 (intermédiaire pauvre) et 3 (intermédiaire riche) sont plus faibles et moins significatifs après instrumentation, au contraire de la classe 4 (riche). L'instrumentation a donc augmenté la pente de la relation entre capital économique et probabilité de scolarisation.

### 7. Valeurs aberrantes et qualité d'ajustement du modèle

Différentes statistiques existent pour détecter les valeurs aberrantes (i.e. les observations très mal prédites) et les points fortement influents dans l'estimation (i.e. les observations modifiant les coefficients de la régression). Il faut combiner ces deux aspects pour détecter des observations, qui seront éventuellement à supprimer ou à traiter de manière particulière. Pour la détection des valeurs aberrantes, nous avons principalement utilisé les résidus standardisés du modèle<sup>13</sup>. On distingue des points fortement aberrants (les valeurs sont supérieures à UOF ou inférieures à LOF<sup>14</sup>), les autres points aberrants étant obtenus à partir de UIF et LIF. Pour la détection des points fortement influents, nous avons utilisé la mesure de Pregibon<sup>15</sup> (équivalent de la dis-

tance de Cook pour les modèles non linéaires). Nous avons retenu qu'un point est fortement influent si le critère est trois fois plus important que la moyenne (cf. tableau 8).

Deux situations méritent d'être analysées : les points à la fois fortement aberrants et influents (24 observations représentant 21 ménages pour 97 observations), ou juste aberrants et influents (166 observations représentant 113 ménages pour 536 observations). Il n'est pas évident de mettre en évidence des spécificités de ces ménages, même si les distances sont plus importantes que la moyenne et que certaines régions sont plus impactées. Mais on peut présumer qu'un certain nombre de points aberrants sont dus à des erreurs de saisie ou de compréhension du questionnaire.

Il nous a semblé nécessaire de supprimer les ménages pour lesquels des observations à la fois fortement aberrantes et influentes ont été détectées. En effet, le test gof est alors nettement amélioré (p-value=0.16 pour 25 classes, 0.41 pour 250 classes, 0.40 pour 500 classes). De plus, le linktest est également légèrement amélioré (p-value associé à la prédiction au carré, 0.014 au lieu de 0.000). Ce test vise à vérifier la pertinence de la spécification. Le principe de ce test est de régresser la variable dépendante sur la valeur prédite et la valeur prédite au carré. S'il n'y a pas d'erreurs de spécifications, le coefficient associé à la valeur prédite au carré ne doit pas être significatif. Ce n'est pas le cas ici. Cette étape nous a également permis de constater l'importance de la prise en compte

## T8 Régression avec ou sans les valeurs aberrantes

	(1) Avec val. aberr.	(2) Sans val. aberr.
Pas d'études	Ref	
Fondamentale 1	0.858 *** (8.68)	0.893 *** (8.89)
Fondamentale 2	1.443 *** (15.64)	1.488 *** (16.05)
Lycée et plus	1.837 *** (13.72)	1.895 *** (13.99)
LireCM	-0.0104 (-0.11)	-0.0185 (-0.20)
Très pauvre	Ref	
Pauvre	0.146 (1.75)	0.165 * (1.96)
Riche	0.320 *** (3.36)	0.325 *** (3.39)
Très Riche	0.700 *** (5.45)	0.721 *** (5.50)
N	12177	12080

Note de lecture : Statistiques de Student entre parenthèses  
Niveaux de significativité : \* : <10% \*\* : <5% \*\*\* : <1%

<sup>13</sup>  $r_i^{std} = \frac{r_i}{\sqrt{1-h_i}}$  avec  $r_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{\sigma}^2(1-h_i)}}$  le résultat observé,  $\hat{y}_i$  la probabilité prédite et  $h_i = \hat{\pi}_i(1 - \hat{\pi}_i)x_i'Var(\hat{\beta})x_i'$

<sup>14</sup> Les bornes sont calculées à partir du premier quartile Q1 et troisième quartile Q3. LOF=Q1-3\*(Q3-Q1), UOF=Q3+3\*(Q3-Q1), LIF=Q1-1.5\*(Q3-Q1), UIF=Q3+1.5\*(Q3-Q1)

<sup>15</sup>  $C_i = \frac{r_i^2 h_i}{1-h_i}$

du plan de sondage. En effet, si on ne supprime que les observations aberrantes et influentes, le test gof conclut à un mauvais ajustement du modèle. Si nous supprimons par contre l'ensemble des ménages présentant un individu aberrant ou influent, le test gof conclut alors à un bon ajustement des données!

Ce qui nous amène à proposer une analyse critique de ce test. En effet, ce test fait l'hypothèse que les observations sont indépendantes entre elles. Ce qui n'est pas le cas dans nos données. Hosmer, Archer et Lemeshow ont en outre proposé, à partir d'exemples dans les biostatistiques, un test valable pour des plans de sondage complexes (Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design, Computational Statistics & Data Analysis, Volume 51, Issue 9 (May 2007)). Nous l'avons mis en place à partir des programmes qu'ils ont implémentés (cf. tableaux 9 et 10).

On remarque qu'à l'exception d'un nombre de classes faible (mais qui correspond au programme des auteurs), l'hypothèse nulle est rejetée. Dans ce cas, même le modèle simple est accepté. Pour des nombres de classes plus élevés, l'interprétation devient plus délicate.

D'autres analyses nous permettent néanmoins de conclure à une bonne qualité discriminatoire du modèle (notamment 75% de classement correct). Nous avons donc retenu pour l'analyse un modèle sans les valeurs aberrantes.

**Conclusion intermédiaire :** Nous avons dû complexifier notre modèle initial, pour obtenir un modèle valide pour l'interprétation. On retrouve le même type d'analyse que pour le modèle simple. Plusieurs remarques sont importantes. Tout d'abord, tenir compte de la structure du plan de sondage est indispensable, certains résultats de tests s'en trouvant modifiés. Cela demeure néanmoins complexe. De plus, nous ne sommes pas à l'abri d'un biais de variables

## T<sub>9</sub> P-value par nombre de classes, GOF normal

	(10)	(25)	(100)	(250)
Modèle de base avec cluster	0.0087	0.0000	0.0000	0.0038
Modèle complet	0.0418	0.0000	0.0000	0.0000
Modèle complet sans les valeurs aberrantes	0.1723	0.1583	0.1939	0.4150

## T<sub>10</sub> P-value par nombre de classes, GOF adapté

	(10)	(25)	(100)	(250)
Modèle de base avec cluster	0.1001	0.0004	0.0000	0.0000
Modèle complet	0.2198	0.0207	0.0000	0.0000
Modèle complet sans les valeurs aberrantes	0.0568	0.0111	0.0103	0.0000

omisées ou d'erreurs d'interprétation des variables (liés à la compréhension du questionnaire), comme semble nous le souligner le link-test. En effet, le contrôle de l'offre d'éducation (à travers la distance à l'école) est faible, on ne tient pas compte des types d'école accessibles, de la taille des classes (saturation). Les dépenses liées à l'éducation, la langue parlée ne sont pas prises en compte directement. Enfin, le caractère endogène de la variable travail mériterait d'être amélioré, la correction ici proposée étant biaisée.

## VI. Interprétations des résultats

### 1. Analyse des effets marginaux

Dans un modèle non linéaire, les signes des paramètres estimés nous informent du sens de variation de la probabilité de scolarisation  $\mathbb{P}(\text{EcoleEC} = 1|x)$ . Par contre, à la différence des modèles linéaires, les effets marginaux  $f(X'\theta)\theta_k$  (i.e. la variation de la probabilité de scolarisation estimée lorsque l'on fait varier une variable explicative d'une unité) ne sont pas constants car ils dépendent de la valeur de tous les coefficients et de toutes les variables, en d'autres termes de la position où l'on se situe pour les évaluer.

Pour pallier ce problème, il existe deux méthodes<sup>16</sup> : le calcul des effets marginaux aux points moyens,

et la moyenne des effets marginaux individuels  $E[f(X'\theta)]\theta_k$ . Si en pratique, les deux méthodes sont utilisées, celles-ci peuvent générer des résultats différents (notamment dus à la non-linéarité du modèle). Selon Bartus [2005], la seconde méthode est la plus pertinente. Nous focaliserons notre analyse sur cette méthode car, dans notre modèle, de nombreuses variables explicatives sont de nature binaire et partitionnées en classes. Utiliser leur moyenne n'aurait que peu de sens, il conviendrait de spécifier des modalités de référence.

La seconde méthode a été proposée par Chamberlain (1982) et consiste donc à calculer la moyenne des effets marginaux individuels. Nous obtenons ainsi la variation en points de pourcentage de la probabilité de scolarisation  $\mathbb{P}(\text{EcoleEC} = 1|x)$ , lorsque l'on fait varier une variable explicative donnée d'une unité (ou de 0 à 1), selon qu'elle soit continue ou discrète.

### Analyse des résultats

Une hausse du capital éducatif au sein du ménage a un effet marginal fortement significatif sur la fréquentation scolaire de l'enfant. En effet, concernant l'école fondamentale premier cycle (primaire en France), même un niveau éducatif faible au sein du ménage accroît consi-

<sup>16</sup> Pour l'étude des effets marginaux des différentes variables explicatives  $E[f(X'\theta)]\theta_k$ , une approche hybride tenant compte des poids aurait également pu être adoptée. Les effets marginaux dépendent en effet conjointement des lois de  $X$  et de  $Y$ . On estime alors le modèle non pondéré (pour calculer  $\theta_k$ ) puis on utilise les poids pour l'estimation du terme  $E[f(X'\theta)]$ .

dérablement les chances de scolarisation de l'enfant. L'accroissement du maximum du niveau d'études au sein du ménage augmente significativement la probabilité de scolarisation de l'enfant : atteindre le collège au sein du ménage implique une chance de scolarisation de l'enfant accrue d'un quart, de près de 32% pour un niveau maximal d'étude au sein du ménage supérieur au lycée. Cependant, de manière remarquable, cet accroissement des chances de scolarisation avec un capital éducatif du ménage croissant sature pour les niveaux éducatifs plus importants : l'effet marginal d'un faible capital éducatif plutôt qu'un capital nul est de 14% alors que celui d'un niveau éducatif du lycée ou plus plutôt que le second cycle n'est que de 7% .

On observe le phénomène inverse pour le capital économique. Un niveau de consommation accrue a une influence significative sur les chances de scolarisation. Cet effet marginal augmente d'autant plus que le proxy du revenu (le niveau de consommation totale) augmente. Ainsi, l'effet d'un niveau de consommation classé comme "pauvre" augmente la probabilité d'aller à l'école de 2,73%. L'effet marginal d'une augmentation de la consommation ("riche") permet une augmentation de la probabilité de scolarisation de 5,43%, alors que celui du niveau supérieur de consommation induit une hausse des chances de scolarisation de l'enfant de 12,23%. Ainsi, marginalement la hausse de scolarisation augmente d'autant plus que le capital économique est important (cf. tableau 11).

De manière remarquable, on note que le capital éducatif semble avoir une influence plus forte que le capital économique dans la décision de scolarisation des enfants<sup>17</sup>.

Par ailleurs, l'effet marginal induit par la variable "travail" est très significatif comme anticipé, avec une diminution des chances de fréquentation scolaire de 15,74%. L'effet des âges extrêmes est également très marqué. Enfin, on note que le genre a un effet marginal significatif, ce que

## T11 Moyenne des effets marginaux individuels

Variable	Effet marginal	Ecart-type
Pas d'études	Ref	
Fondamentale 1	14,55 ***	(1,58)
Fondamentale 2	25,33***	(1,45)
Lycée et plus	31,96 *	(1,82)
Très pauvre	Ref	
Pauvre	2,73 **	(1,39)
Riche	5,43 ***	(1,60)
Très Riche	12,23 ***	(2,16)
Sexe (femme)	-6,45***	(0,75)
Age 6	-34,93 ***	( 1,05)
Age 15	-12,73 ***	(1,21)
Milieu	-3,60 **	( 1,73)
Travail	-15,74 ***	(3,28)

Note de lecture : Niveaux de significativité : \* : 10% \*\* : 5% \*\*\* : 1%

nous allons étudier de manière plus détaillée dans la partie suivante.

### 2. Etude de la stabilité des paramètres entre genres

Un argument souvent avancé dans la littérature suggère l'existence de différences dans les déterminants de la fréquentation scolaire d'un enfant, selon que celui-ci soit un garçon, ou une fille. On s'intéresse dans cette section à ce problème, et plus particulièrement aux manières de spécifier cette caractéristique du genre.

Notre modèle initial se contentait d'introduire une indicatrice pour capter l'effet engendré par le fait d'être une fille. Pour rappel, dans la version classique de notre modèle, cet effet était négatif et significatif, traduisant la fréquentation plus faible des filles au système scolaire. On envisage ainsi dans un second temps, de tester la pertinence d'un modèle opérant une séparation plus marquée. Il consisterait à séparer totalement les deux échantillons, en partant de l'hypothèse que les ressorts de l'éducation ne sont pas les mêmes entre les deux sexes. Nous testerons tout d'abord cette hypothèse, avant d'analyser ce type de modèle.

#### Test de stabilité de paramètres

Ce type de test est souvent développé dans le cadre de séries tem-

porelles, pour tester l'existence d'un choc structurel à une date  $t$ , modifiant les paramètres estimés. Ce test, appelé communément test de Chow, va ici être adapté pour s'appliquer à notre situation. En effet, alors que le test de Chow compare la performance d'un modèle contraint par l'hypothèse de stabilité à celle d'un modèle non contraint, nous utiliserons un test d'égalité des paramètres uniquement sur ceux correspondant à nos variables d'intérêt. Etant donné le nombre de variables de contrôles dans le modèle (environ quarante), une égalité parfaite de tous les coefficients est une hypothèse trop forte à tester dont le rejet semble évident. On se limite donc à un test d'égalité des six coefficients d'intérêts attachés aux variables de consommation et d'éducation du chef de ménage.

On pose donc  $H_0 : \beta_1 = \beta_2$ . Un test de Wald est construit après avoir extrait des deux régressions, les vecteurs de paramètres, et les matrices de variance associées. On rappelle que sous l'hypothèse  $H_0$ , la statistique de Wald suit une loi du Chi-deux à  $K$  degré de liberté,  $K$  étant le nombre de contraintes fixées. On obtient ainsi

$$W = \hat{\beta}' R' [R\hat{V}(\hat{\beta})R']^{-1} R\hat{\beta} = 4.09$$

avec  $\beta = (\beta_1, \beta_2)$  et  $R = (IK - IK)$

<sup>17</sup> Les résultats des modèles avec et sans les variables d'intérêt corroborent ce phénomène. Le pseudo-R2 du modèle complet est de 0.2780. Sans le capital éducatif (le niveau d'éducation), il tombe à 0.2344 alors que sans le capital économique (la consommation totale), il est de 0.2744



Sous l'hypothèse  $H_0$ , la statistique  $W$  suit une loi du Chi-deux à  $K=6$  degrés de liberté,  $K$  étant le nombre de paramètres, et donc de contraintes imposées. Or, le quantile à 95% de cette loi est à 12,59, l'hypothèse  $H_0$  d'égalité des paramètres entre les deux régressions est donc acceptée.

Concrètement, cela signifie que l'introduction d'une indicatrice est suffisante pour prendre en compte les effets du sexe sur le rôle des capitaux économiques et culturels. L'influence des déterminants est donc statistiquement non différente selon que l'enfant soit une fille ou un garçon.

### Comparaison des régressions

On présente dans le tableau 12 les résultats des régressions séparées suivant le genre. Bien que ces différences soient nulles d'un point de vue statistique, leur étude est néanmoins intéressante.

Concernant l'effet du niveau d'étude au sein du ménage, les différences entre garçons et filles ne sont pas très importantes. Néanmoins, il est remarquable de noter que l'éducation des membres du ménage profite plus aux filles qu'aux garçons. Ceci est le cas pour l'effet sur la scolarisation d'un parent ayant été au primaire et au collège. En revanche, l'effet marginal du secondaire et du supérieur par rapport au collège semble beaucoup plus faible pour les filles.

Ainsi, on comprend facilement que les premiers niveaux d'étude d'un parent favorise l'éducation des filles à travers un dépassement des normes culturelles. Avoir été à l'école offre sans aucun doute un regard plus avancé sur la place de la femme dans la société moderne. Néanmoins, cet effet diminue de plus en plus avec l'éducation des parents, les années d'études secondaires et supérieures n'ayant pas d'effet propre par rapport au primaire. Ainsi, les parents très éduqués n'auront plus de discrimination entre garçons et filles dans leur choix de scolarisation.

## T12 Régression séparée par genre

	(1) Garçons	(2) Filles
Pas d'études	Ref	
Fondamentale 1	0.780*** (0.094)	0.913*** (0.105)
Fondamentale 2	1.362*** (0.084)	1.420*** (0.088)
Lycée et plus	1.819*** (0.128)	1.732*** (0.130)
Très pauvre	Ref	
Pauvre	0.146 (0.076)	0.153 (0.085)
Riche	0.257** (0.089)	0.399*** (0.096)
Très Riche	0.564*** (0.133)	0.753*** (0.133)

Note de lecture : Ecart-type entre parenthèses  
Niveaux de significativité : \* : 10% \*\* : 5% \*\*\* : 1%

Encore une fois, il faut modérer nos explications par le fait que ces écarts ne sont pas significatifs d'un point de vue statistique. Bien que commentés, ces différences ne sont donc pas nécessairement validées.

## Conclusion

Dans un pays en voie de développement tel que le Mali, la scolarisation des enfants est considérée comme un enjeu de premier plan. Si les efforts publics, les initiatives privées, et le support des ONG vont largement dans le sens d'une augmentation massive de l'offre d'éducation de premier et second cycle depuis les vingt dernières années, il apparaît important de s'intéresser aux déterminants de la demande en éducation dans ce pays.

En d'autres termes, dans quelle mesure et par quel canal, les capitaux économique et éducatif des ménages permettent-ils de déterminer la scolarisation ou non des enfants de 6 à 15 ans?

Pour répondre à cette question, nous nous sommes appuyés sur l'enquête ELIM (2006), menée sur un large échantillon de ménages représentatifs de la population entière du Mali. De par les multiples domaines qu'elle couvre, elle nous a permis d'identifier, mesurer et contrôler les effets du capital économique et éducatif au sein du ménage dans la scolarisation des enfants.

Au regard d'un contexte très spécifique au Mali et plus généralement des pays francophones d'Afrique subsaharienne, nos analyses descriptives et économétriques ont mis en évidence l'importance a priori intuitive de nos deux dimensions explicatives, niveau de consommation totale et d'éducation du ménage, dans la scolarisation des enfants. Fait notable dans l'analyse descriptive, pour un niveau inférieur à six années d'étude, savoir lire et écrire n'est pas une évidence. Pourtant, la variable *LireEcrire* n'apparaît pas significative dans les régressions économétriques.

L'analyse des effets marginaux de notre modèle économétrique indique qu'un accroissement des capitaux économique et éducatif du ménage permet une fréquentation scolaire accrue des enfants. En outre, le capital éducatif a un effet prépondérant au regard de celui du capital économique. De plus, les chances de scolarisation d'un enfant

sont d'autant plus grandes que le niveau de consommation par tête total du ménage est important. A l'inverse, on observe concernant le capital éducatif maximal du ménage un phénomène de saturation pour les niveaux d'éducation supérieurs.

Par ailleurs, de nombreuses inégalités dans la scolarisation des enfants au Mali sont constatées que ce soit en termes d'opposition entre milieu (urbain,rural) ou entre sexe. Si les différences de scolarisation entre sexe est sans contexte, nous avons cependant montré que l'influence des déterminants économique et éducatif du ménage sur la scolarisation des enfants n'est pas différente selon le genre.

D'un point de vue technique, l'analyse des déterminants propres aux ménages dans la scolarisation de leurs enfants a soulevé de nombreuses questions. La majeure partie de l'analyse économétrique du modèle dichotomique a été menée sur un modèle logistique, même si certains tests se sont appuyés sur un modèle probit.

En premier lieu, notre réflexion a tenu compte du plan de sondage, tant pour savoir si le modèle devait être pondéré ou non que pour la correction de la matrice de variance. Nous avons également considéré l'éventualité de biais d'endogénéité dans notre modèle initiale. L'analyse menée a en outre montré que si la théorie économique suggère la possibilité d'une "fécondité endogène" (Becker), cela n'est pas le cas dans notre population. Cependant, l'endogénéité provenant de la variable *travail* a été montrée. La correction apportée demeure néanmoins biaisée. Enfin, des valeurs aberrantes ont été détectées, qui ont été supprimées de notre modèle final notamment de par leurs effets sur les tests statistiques.

Ainsi, il a été établi que la fréquentation scolaire des enfants en âge d'être scolarisé au Mali est largement déterminée par le milieu familial d'appartenance, plus précisément par le pouvoir d'achat du ménage et sa connaissance (et reconnaissance) du système éducatif malien. Un enseignement certain

de notre étude est que le capital éducatif est largement prépondérant sur le capital économique dans la décision de scolarisation des enfants. L'implication de ce résultat en termes de politiques économiques serait une concentration des pouvoirs publics sur l'offre d'éducation, l'alphabétisation des adultes, des campagnes de sensibilisation et d'informations quant au système scolaire malien, bien plus que sur la mise en place de transferts monétaires vers les ménages.

## Bibliographie

### Articles d'économie empirique

**Duflo, Esther, Pascaline Dupas, Michael Kremer** (2008), "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," mimeo MIT Abdul Latif Jameel Poverty Action Lab

**Kremer, Michael, Edward Miguel, and Rebecca Thornton** (2008), "Incentives to Learn," Review of Economics and Statistics (forthcoming)

**Nguyen, Trang** (2008), "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar," mimeo MIT

**Behrman J.R., Wolfe B.L.** (1984) Who is schooled in developing countries ? The roles of income, parental schooling, sex, residence and family size. *Economics of Education Review*. Vol. 3. No. 3. pp. 231-241, 1984

**Okumu, Ibrahim M., Naka jjo, Alex and Isoke, Doreen** (2008). Socioeconomic determinants of primary school dropout: the logistic model analysis. Economic Policy Research Center, Makerere University

**Bowman M.J.** (1984). An Integrated Framework for Analysis of the Spread of Schooling in Less Developed Countries. *Comparative Education Review*, Vol. 28, No. 4 (Nov., 1984), pp. 563-583

**Buchmann C.** (2000). Family Structure, Parental Perceptions, and Child Labor in Kenya: What Factors Determine Who is Enrolled in School? *Social Forces*, Vol. 78, No. 4 (Jun., 2000), pp. 1349-1378

**Hughes R.** (1991). Examining the Roots of Educational Demand: The Case Supporting Rural Agrarian Development. *World Development*, Vol. 19, No. 2/3, pp. 213--223

**Gajendra Man Shrestha, Sri Ram Lamichhane, Bijaya Kumar Thapa, Roshan Chitrakar, Michael Useem, John P. Comings** (1986). Determinants of Educational Participation in Rural Nepal. *Comparative Education Review*, Vol. 30, No. 4 (Nov., 1986), pp. 508-522

**Kabore I., Lairez T., Pilon M.** (2002). Genre et scolarisation au Burkina Faso : enseignements d'une approche statistique.

**Cynthia B. Lloyd and Ann K. Blanc** (1996). Children's Schooling in sub-Saharan Africa: The Role of Fathers, Mothers, and Others. *Population and Development Review*, Vol. 22, No. 2 (Jun., 1996), pp. 265-298

**Martin Ravallion and Quentin Wodon** (2000). Does child labour displace schooling ? Evidence on behavioural responses to an enrollment subsidy. *The Economic Journal*, 110 (March), C158±C175

**Odaga, A. and Heneveld, W.** (1995) Girls and schools in sub-Saharan Africa: From analysis to action World Bank Publications

### Articles statistiques

**Archer K.J., Lemeshow S., Hosmer D.W.** (2007). Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design. *Computational Statistics - Data Analysis* 51 (2007) 4450-4464

**D'Haultfoeuille Xavier, Davezies Laurent** (2009). Faut-il pondérer? ... ou l'éternelle question de l'économètre confronté à un problème de sondage, Note préliminaire

**Hosmer David W. Jr, Lemeshow Stanley** (2000). Applied logistic regression. Wiley series in probability and statistics.

**Wooldridge Jeffrey M.** (2000). Econometric Analysis of Cross Section and Panel Data. MIT Press

### Articles de théorie économique

**Becker G.S., Murphy K.M., Tamura R.** (1990) Human capital, fertility, and economic growth. *Journal of Political Economy*, Vol. 98, No. 5, Part 2: The Problem of Development: A Conference of the Institute for the Study of Free Enterprise Systems, pp. S12-S37

**Werdelin I.** (1970). A School Enrolment Model. *International Review of Education / Internationale Zeitschrift für Erziehungswissenschaft / Revue Internationale de l'Education*, Vol. 16, No. 2 (1970), pp. 192-214

### Articles contextuels

**IMF** (2008). Poverty Reduction Strategy Paper. <http://www.imf.org/external/NP/prsp/prsp.asp#M>

**Konate M.K., Guèye M., Nseka Vita T.** (2003). Scolarisation des enfants au Mali selon le profil des ménages et étude de leur maintien à l'école. Background paper prepared for the Education for All Global Monitoring Report 2003/4 Gender and Education for All: The Leap to Equality

**Konate M.M., Tamboura P.** (1999). Etude prospective/bilan de l'éducation en Afrique, le cas du Mali. Association pour le développement de l'éducation en Afrique (ADEA)

**Lange M-F.** (2001). L'évolution des inégalités d'accès à l'instruction en Afrique depuis 1960. UEPA/UAPS, INED, ENSEA, IFORD Abidjan, 16-21 juillet 2001

**Opheim M.** (2000). Les filles et l'école au Mali. Nordic Journal of African Studies 9(3): 152-171

### Divers, sociologie de la mobilité sociale

**Thélot C.** (1982). Tel Père, tel Fils? Dunod

**Thélot C., Vallet L-A.** (2000). La réduction des inégalités sociales devant l'école depuis le début du siècle. Economie et statistique, n°334, 4, p. 3-32

**Swift, A.** (2004). Would perfect mobility be perfect ?, European Sociological Review, 20, 1-11  
Guide de l'enquêteur, rapport d'analyse et questionnaires de l'enquête ELIM 2006.

## Annexes

### 1. Caractéristiques générales des variables de l'échantillon

L'enquête ELIM a utilisé un plan de sondage à 2 degrés, dont l'unité primaire, la grappe, est la Section d'Enumération (SE) telle que définie au Recensement Général de la Population et de l'Habitat (RGPH) du Mali de 1998 et l'unité secondaire, le ménage. Pour améliorer la précision, il a également été tenu compte d'une stratification au niveau de la région (urbaine et rurale) et des ménages (selon le type de budget). Cette enquête a été coordonnée avec l'échantillon des unités primaires de l'Enquête Malienne d'Evaluation de la Pauvreté 2001. L'échantillon tiré a porté sur 750 unités primaires et 4 500 ménages (à raison de 6 ménages par unité primaire tirée).

Seule une grappe (unité primaire) n'a pu être interrogée. Nous avons utilisé des variables relatives à l'éducation, au travail, aux dépenses de consommation et d'accès aux infrastructures. Les taux de non-réponse sont très faibles pour ces variables. Ils sont en général inférieurs à 0.5% sauf pour les variables d'accès aux infrastructures pour lesquelles ils représentent un peu plus de 1%. De plus, pour ces dernières, la non-réponse portait souvent sur l'ensemble des ménages d'une même grappe. Pour les variables d'accès aux infrastructures, nous avons donc conservé une modalité 'non-déclaration'. Pour les autres variables et pour la mise en place du modèle, la modalité 'non-déclaration' a été regroupée avec d'autres modalités de manière déterministe.

Les tableaux 13 et 14 décrivent les principales variables descriptives et continues de l'enquête. Les statistiques brutes et pondérées (par les poids de l'enquête) sont calculées.

**Tableau 13 : Caractéristiques des principales variables continues**

	Moyenne non pondérée	Moyenne pondérée
Age du chef de ménage	50,66 (13,15)	50,09 (13,10)
Niveau d'étude du chef de ménage (nombre d'années)	2,39 (4,32)	1,82 (3,79)
Niveau d'étude du conjoint ou des conjoints (nombre d'années)	1,17 (2, 93)	0,97 (2,64)
Nombre d'enfants (fils ou fille)	4,72 (2,96)	4,58 (2,85)
Taille du ménage	10,10 (5,83)	9,63 (5,50)
Dépenses totales par tête (milliers CFA)	185,55 (174,23)	183,1 (159,96)
Dépenses alimentaires par tête (milliers CFA)	93,87 (57,38)	94,79 (56,92)
Proportion des dépenses alimentaires sur les dépenses totales (%)	56,57 (15,05)	56,80 (14,58)
Distance à l'école primaire (km)	1,81 (6,48)	1,69 (6,40)
Distance au collège (km)	6,27 (11,41)	6,27 (11,58)

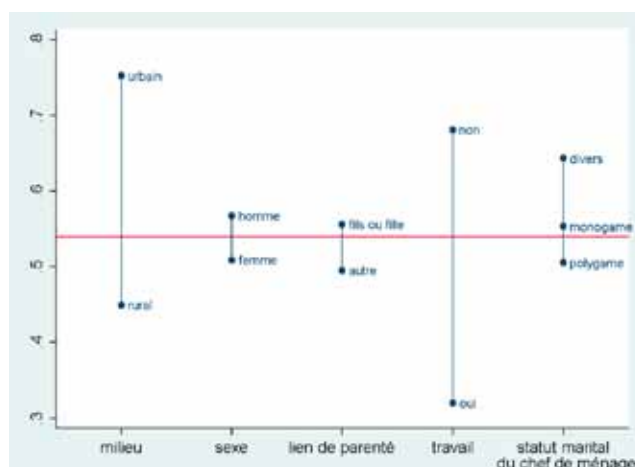


**Tableau 14 : Caractéristiques des principales variables discrètes**

Variable	Modalités	Fréquence	Distribution non pondérée (%)	Distribution pondérée (%)
Milieu	Urbain	3764	30,3	30,1
	Rural	8648	69,7	69,9
Sexe de l'enfant	Homme	6695	53,9	52,6
	Femme	5717	46,1	47,4
Sexe du chef de ménage	Homme	3478	93,9	92,8
	Femme	227	6,1	7,2
Statut marital du ménage du chef de ménage	Monogamme	2288	61,8	61,7
	Polygame	1163	31,4	30,1
	Autres situations	254	6,9	8,2
Travail de l'enfant	Oui	4860	39,2	38,6
	non	7552	60,8	61,4
Scolarisation de l'enfant	Oui	6075	48,9	48,5
	A abandonné	345	2,8	2,8
	N'a jamais fréquenté l'école	5992	48,3	48,7
Niveau de lecture et d'écriture du chef de ménage	Sait lire et écrire	1125	30,4	26,5
	Analphabète	2580	69,6	73,5
Catégorie socioprofessionnelle du chef de ménage	Cadre ou patron, employeur	299	8,1	5,8
	Employé, ouvrier	307	8,3	8,1
	Pour compte propre	2506	67,6	71,3
	Manoeuvre, aide familial ou apprenti	76	2,1	1,8
	Sans emploi	517	14,0	13,1
Indice de pauvreté	Pauvre	1637	44,2	43,4
	Non pauvre	2068	55,8	56,6

## 2. Fréquentation scolaire et variables de contrôle

**Figure 10 : Fréquentation scolaire et principales variables de contrôle**



**Figure 11 : Fréquentation scolaire et distance aux infrastructures**

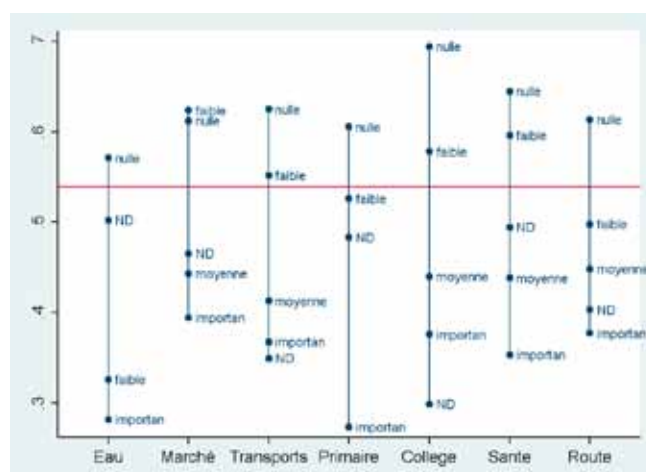


Figure 12 : Fréquentation scolaire et CSP

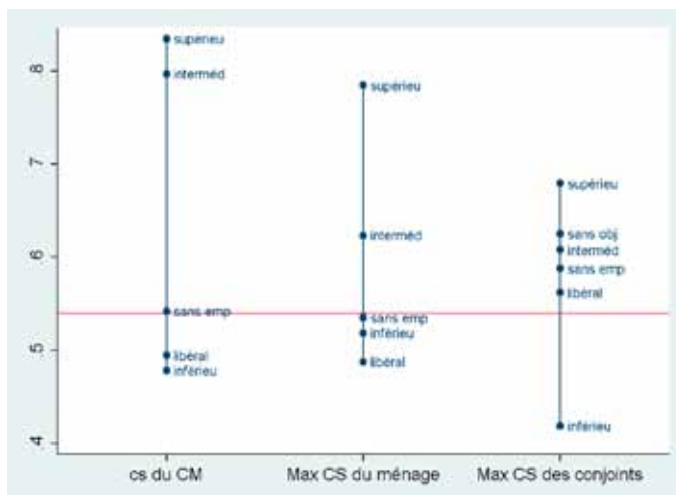
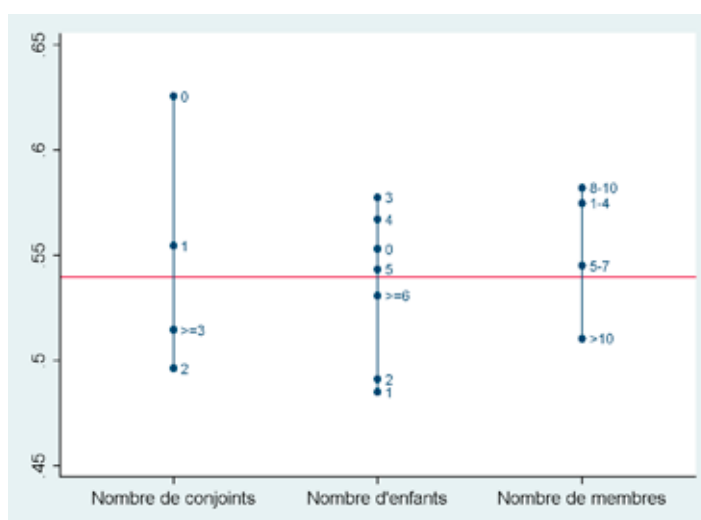


Figure 13 : Fréquentation scolaire et taille du ménage (nombre d'enfants)



### 3. Détails théoriques sur l'hétéroscédasticité

Pour voir pourquoi les coefficients d'un modèle à effets fixes ne peuvent être estimés, réexprimons notre modèle

$$y_{j,i}^* = x'_{i,j} \beta_1 + z'_j \beta_2 + v_j + e_{i,j} \quad (3)$$

Avec  $y_{j,i}^*$  la variable latente associée à la variable scolarisation de l'enfant  $i$  du ménage  $j$ ,  $z_j$  le vecteur de variables ménage commun à tous les enfants du ménage,  $x_{i,j}$  un vecteur de variables explicatives propres à chaque enfant,  $v_j$  un effet fixe inobservable commun à l'ensemble des enfants du ménage  $j$  et corrélé avec  $(x_{i,j}, z_j)$  et enfin  $e_{i,j}$  le «vrai» résidu du modèle au sens où il respecte l'hypothèse d'homoscédasticité.

Exprimons la vraisemblance du deuxième enfant du ménage  $j$  conditionnellement à  $x_{i,j}, z_j, v_j$ , et  $n_j = \sum_{i=1}^{n_j} y_{i,j}$

Supposons pour simplifier que le ménage contient deux enfants ( $n_j = 2$ ).  $n_j$  étant le nombre d'enfant dans le foyer,  $P(y_{i2} | x_{i,j}, z_j, v_j, n_j)$  est donc totalement déterminé pour  $n_j = 0$  ou  $n_j = 2$ . Les ménages dans ce cas ne contribuent donc pas à la vraisemblance. Prenons donc le cas  $n_j = 1$ .

$$P(y_{i2} | x_{i,j}, z_j, v_j, n_j = 1)$$

$$= \frac{P(y_{i2} = 1, n_j = 1 | x_{i,j}, z_j, v_j)}{P(n_j = 1 | x_{i,j}, z_j, v_j)}$$

$$\begin{aligned}
&= \frac{P(y_{i2}=1, | x_{i,j}, z_j, v_j) P(y_{i1}=1, | x_{i,j}, z_j, v_j)}{P(y_{i2}=1, y_{i1}=0 | x_{i,j}, z_j) + P(y_{i2}=0, y_{i1}=1 | x_{i,j}, z_j)} \\
&= \frac{F(x'_{j,2} \beta_1 + z'_j \beta_2 + v_j) [1 - F(x'_{j,1} \beta_1 + z'_j \beta_2 + v_j)]}{\{F(x'_{j,1} \beta_1 + z'_j \beta_2 + v_j) [1 - F(x'_{j,2} \beta_1 + z'_j \beta_2 + v_j)] + F(x'_{j,2} \beta_1 + z'_j \beta_2 + v_j) [1 - F(x'_{j,1} \beta_1 + z'_j \beta_2 + v_j)]\}} \\
&= F((x'_{j,2} \beta_1 + z'_j \beta_2 + v_j) - (x'_{j,1} \beta_1 + z'_j \beta_2 + v_j)) \\
&= F((x'_{j,2} - x'_{j,1}) \beta_1)
\end{aligned}$$

De la même manière on obtient

$$P(y_{i2} = 1 | x_{i,j}, z_j, v_j, n_j = 1) = 1 - F(x'_{j,2} - x'_{j,1}) \beta_1$$

**La vraisemblance ne dépendant plus de  $\beta_2$ , il ne peut être identifié.**

Une autre solution est d'estimer un logit avec effet aléatoire. La contribution à la vraisemblance au sein d'un ménage  $j$  est

$$\begin{aligned}
&P(y_{1j}, y_{ne_j} | x_{ij}) \\
&= \int P(y_{1j}, y_{ne_j}, v_j | x_{ij}) dv_j \\
&= \int P(y_{1j}, y_{ne_j} | x_{ij}, v_j) dG(v_j | x_{ij}).
\end{aligned} \tag{4}$$

Avec  $ne_j$  le nombre d'enfants en âge d'aller à l'école dans le ménage  $j$ ,  $G(\cdot | x_{i,j})$  la loi de  $v_j$  conditionnellement à  $x_{i,j}$ . L'effet aléatoire étant supposé non corrélé avec  $x_{i,j}$ , on a  $G(v_j | x_{i,j}) = G(v_j)$ , avec  $G(\cdot)$  la distribution de  $v_j$ <sup>18</sup>, supposée identique dans l'ensemble des ménages. A partir de (4) on obtient l'expression de la contribution d'un ménage à la vraisemblance

$$L(y_{1j}, y_{ne_j} | x_{ij}) = \int \prod_{i=1}^{ne_j} F(x'_{ij} \beta + u_j)^{y_{ij}} [1 - F(x'_{ij} \beta + u_j)]^{1-y_{ij}} dG(v_j)$$

La vraisemblance du modèle est

$$\prod_{j=1}^J L(y_{1j}, y_{ne_j} | x_{ij})$$

Avec  $J$  le nombre de ménages observés.

<sup>18</sup> Cette distribution étant inconnue, elle implique d'introduire une nouvelle hypothèse dans le modèle. Par défaut STATA suppose une distribution normale.

## Liste des publications AFRILUX en 2008-2009

TENIKUE Michel, VERHEYDEN Bertrand. Birth Order, Child Labor and Schooling: Theory and Evidence from Cameroon. CEPS/INSTEAD, 2008, **Afrilux** n°5, 16 p.

DOUMBIA-GAKOU Assa, KUEPIE Mathias. Les déterminants de l'insertion des femmes sur le marché du travail au Mali. CEPS/INSTEAD, 2008, **Afrilux** n°4, 16 p.

COULIBALY Ishaga, KEITA Balla, KUEPIE Mathias. Les déterminants du recours thérapeutique au Mali : entre facteurs socioculturels, économiques et d'accessibilité géographique. CEPS/INSTEAD, 2008, **Afrilux** n°3, 16 p.

DABITAO Kassim, KUEPIE Mathias. Les déterminants de l'insertion sur le marché du travail au Mali. CEPS/INSTEAD, 2008, **Afrilux** n°2, 20 p.

KUEPIE Mathias, DOUMBA-GAKOU Assa. Les mesures de la pauvreté objective au Mali : à méthodes différentes, résultats différents ? CEPS/INSTEAD, 2008, **Afrilux** n°1, 16 p.

# AFRILUX

### **CEPS/INSTEAD**

B.P. 48

L-4501 Differdange

Tél. : 58 58 55-801

e-mail : [isabelle.bouvy@ceps.lu](mailto:isabelle.bouvy@ceps.lu)

<http://www.ceps.lu>