



---

## **La fonction de revenu de Mincer : fondements théoriques et analyse critique**

**Yassir CHAKIB et Siham LAMARI**

FSJES AIN SEBAA, LARMIG

---

### **Résumé**

Pourquoi certaines personnes gagnent trois fois ou quatre fois plus que d'autres ? Selon la théorie du capital humain, les disparités salariales s'expliquent par les différences d'investissement en capital humain. Théoriquement, les personnes mieux payées sont dotées d'un capital humain plus important accumulé avec l'investissement. L'éducation est l'une des principales formes d'investissement en capital humain. Cela implique qu'il existe une relation entre les revenus et le niveau éducatif. Mincer défend théoriquement et présente un modèle empirique qui permet de tester la relation revenu-éducation sur les données observées. Le modèle permet également sous certaines hypothèses d'estimer le taux de rendement d'une année d'études. La fonction de revenu de Mincer a connu un succès fulgurant. Une large littérature s'est développée en rapport direct avec le modèle. L'équation de Mincer a été largement utilisée dans plusieurs études et appliquée sur une multitude de bases de données relatives aux différents pays. De manière générale, les résultats des études réalisées indiquent l'existence d'un effet du niveau éducatif sur le revenu. Plusieurs critiques constructives ont été dressées à l'égard du modèle. Dans le but de les dépasser, des solutions ont été proposées et des améliorations ont été apportées à l'équation des revenus. Le présent travail est une synthèse d'une revue de littérature sélective des travaux réalisés en rapport avec la fonction du revenu de Mincer. Nous présentons les fondements théoriques de base du modèle ses mérites et ses limites.

### **Abstract**

Why do some people earn three or four times more than others? According to the human capital theory, wage gaps are explained by the differences in investment on human capital. Theoretically, the highest-paid persons are the ones who are endowed with more human capital accumulated with investment. Education is one of the main forms of investment in human capital. This implies that there is a relationship between income and educational level. Mincer theoretically defends and presents an empirical model that allows the relationship between income and education to be tested on observed data. The model also makes it possible, under certain assumptions, to estimate the rate of return within a year of study. Mincer's income feature has been wildly successful. A large literature has been directly developed related to the model. Mincer's equation has been widely used in several studies and applied to a multitude of databases related to different countries. In general, the results of the conducting studies, indicate the existence of an effect of the level of education on income. Several constructive criticisms have been leveled at the model. To overcome them, solutions have been proposed and improvements have been made to the revenue equation. This work is a synthesis of a selective literature review of the works done in relation to Mincer's income function. We present the basic theoretical foundations of the model's merits and limitations.

**Keywords:** Gain Function ; Mincer Model ; Wage Gaps ; Rate Of Return ; Investment In Education ; Literature Review.

**Digital Object Identifier (DOI) :** <https://doi.org/10.5281/zenodo.7262273>

---

## 1 Introduction :

Investir en capital humain est aujourd'hui l'une des grandes préoccupations des pouvoirs publics au niveau mondial. L'accumulation du capital humain engendrerait des gains de productivité favorables à la croissance et à l'emploi.

Le ralentissement de la croissance économique que la plupart des pays occidentaux ont connu depuis 1973 et la montée du chômage ont ouvert une porte de recherche sur la relation entre le capital humain et la productivité.

La main d'œuvre a toujours représenté un avantage concurrentiel pour les entreprises ou même pour les pays. La notion de spécialisation de Smith ou même le concept des avantages comparatifs de Ricardo peuvent y être liés. Les origines de la théorie moderne du capital humain remontent toutefois à la fin des années 1950 avec les travaux de Theodore Schultz (Prix Nobel 1979), Gary Becker (Prix Nobel 1992) et Jacob Mincer qui ont fondé la théorie du capital humain.

La théorie du capital humain défend l'idée que l'éducation est un investissement qui coûte en termes de temps et en termes d'argent. Un individu rationnel ne consentira à supporter les coûts occasionnés par l'éducation que si cet investissement lui rapporte un surplus de gain dans le futur. Si un individu sans qualifications gagne toutes choses égales par ailleurs un revenu fixe de l'ordre de  $T$  tout au long de sa vie professionnelle ; une personne qui fait le choix d'investir dans son éducation gagne durant les années de formation  $T^- < T$  puisque elle doit supporter les coûts de l'investissement mais à la fin de sa formation son revenu grimpe jusqu'à  $T^+ > T$  [1]. Il s'agit évidemment d'un schéma caricatural puisque l'accumulation du capital humain est graduelle et par conséquent le revenu augmente au fur et à mesure. Cela suppose qu'il existe une relation entre ce qu'apprend l'individu durant les cours et son revenu.

Sur un marché de concurrence pure et parfaite, le salaire rémunère la productivité. Par conséquent, un effet positif de l'éducation sur le revenu signifie que le savoir et le savoir-faire transmis durant la formation impacte la productivité de l'individu. Il s'agit effectivement de la chaîne de causalité défendue par la théorie du capital humain « éducation-productivité-salaire ». Sur le plan empirique il n'est pas aisé de mesurer directement la productivité. En supposant le caractère concurrentiel du marché de travail, le revenu devient lui-même une mesure de productivité. Tester empiriquement la relation éducation-revenu est alors suffisant.

Le modèle de Mincer [2] apporte un fondement théorique à la relation directe entre éducation et revenu. Il devient par la suite un modèle économétrique largement utilisé, et par conséquent se développe autour de la fonction de gain de Mincer une vaste littérature. Le but n'est pas seulement de valider empiriquement l'existence d'une relation entre revenu et éducation mais également de mesurer le rendement de l'investissement en capital humain par pays, par région, par genre et par secteur.

Dans ce papier nous proposons une revue de littérature sélective dans le but d'exposer les fondements du modèle, ses mérites ainsi que ses limites. Cet article se compose de deux grandes parties. Dans un premier temps, nous allons exposer le fondement théorique, nous abordons la validité empirique du modèle, les résultats d'un certain nombre de tests empiriques, nous présentons aussi une analyse comparative par pays, par région, par genre et par secteur, avant de terminer la première partie par un focus spécial sur le cas marocain.

Dans la deuxième partie nous traitons les limites du modèle empirique dans la mesure où nous allons évoquer le problème d'endogénéité et de sélectivité, la difficulté de mesurer le volume de l'investissement ainsi que le problème posé par le revenu mixte. Nous présentons aussi les résultats d'études remettant en cause la linéarité du modèle log-revenu-éducation. A la fin nous allons terminer par présenter les solutions pour corriger voire dépasser ces limites.

## 2 Fonction de gain cadre théorique et validité empirique

### 2.1 La relation éducation-revenu : Cadre théorique

L'idée de base de la relation études-revenu est que l'éducation est un investissement qui occasionne des coûts mais rapporte en contrepartie un surplus de revenu. Les frais d'inscription, la fourniture scolaire, les frais de transport, le loyer sont des exemples de postes de charges entrant dans la composition du coût direct occasionné par l'éducation. Même dans le cas où l'Etat assure un enseignement gratuit, offre la fourniture scolaire sans contrepartie et construit des pensionnats pour les personnes éloignées, l'éducation coûte toujours le temps sacrifié. Afin de poursuivre leurs études les élèves et les étudiants sacrifient le temps qu'ils auraient pu dédier à un travail rémunéré. A ce titre, l'éducation est alors un investissement. Une personne rationnelle ne consentira à investir dans sa formation que si celle-ci lui rapporte un supplément de gain futur (ici on néglige la valeur intrinsèque de la formation). Ainsi, les coûts supportés doivent être corrélés avec le revenu attendu. Si on fait le choix de négliger les coûts directs, le volume de l'investissement peut alors être mesuré par le nombre d'années d'étude, d'où la relation 'années d'étude-revenu'. Cette relation a été développée et défendue théoriquement et empiriquement par Mincer [2]. Dans ce qui suit nous présentons son modèle qui a connu un succès fulgurant au sein de la communauté scientifique.

#### 2.1.1 Le modèle du rendement de l'investissement scolaire

Si on pose :

- $n$  : durée de vie professionnelle plus le nombre d'années de scolarisation. Pour les personnes non éduquées  $n$  correspond à l'étendue de leur vie professionnelle
- $Y_K$  : est le revenu annuel reçu par un individu après  $s$  années d'étude.
- $V_K$ : La valeur présente totale des revenus reçus tout au long de la vie à la date du début de sa scolarité.
- $r$ : Taux d'actualisation.
- $t = 0, 1, 2, 3 \dots n$  temps en années.

Alors on peut écrire

$$V_K = Y_K \int_0^n e^{-rt} dt = \frac{Y_K(e^{-rK} - e^{-rn})}{r}$$

Si maintenant l'individu au lieu d'étudier  $K$  années d'étude, il fait le choix d'étudier  $K-D$  années d'étude, la formule devient alors :

$$V_{K-D} = \frac{Y_{s-d}(e^{-r(K-D)} - e^{-rn})}{r}$$

On remarque que la date de retraite est maintenue fixe

On égalise les deux valeurs présentes puis on calcule le ratio  $\frac{Y_K}{Y_{K-D}}$  noté  $K_{K,K-D}$  qui détermine la condition qui fait que l'individu est indifférent entre les deux options :

$$K_{K,K-D} = \frac{Y_K}{Y_{K-D}} = \frac{e^{-r(K-D)} - e^{-rn}}{(e^{-rK} - e^{-rn})} = \frac{e^{r(n+D-K)} - 1}{e^{r(n-K)} - 1}$$

On remarque que le ratio est supérieur à 1. Il s'accroît avec le taux d'actualisation et décroît suite à un rallongement de la vie professionnelle. Cela signifie que les attentes salariales d'un individu rationnel sont d'autant plus élevées que la durée de formation est longue, que le coût du temps est élevé et que la durée de récupération de l'investissement est courte.

Pour plus de simplicité on fixe  $n$  la durée de la vie professionnelle. La date de retraite devient alors variable. En recalculé de nouveau et on obtient :

$$V_K = Y_K \int_K^{n+K} e^{-rt} dt = \frac{Y_K}{r} e^{-rK} (1 - e^{-rn})$$

$$V_{K-D} = Y_{K-D} \int_{K-D}^{n+K-D} e^{-rt} dt = \frac{Y_{K-D}}{r} e^{-r(K-D)} (1 - e^{-rn});$$

Le ratio  $K_{K,K-D}$  se présente alors comme suit :

$$K_{K,K-D} = \frac{Y_K}{Y_{K-D}} = \frac{e^{-r(K-D)}}{(e^{-rK})} = e^{rD}$$

Pour  $K - D = 0$  on a

$$K_{K,0} = \frac{Y_K}{Y_0} = K_K = e^{rK}$$

Il s'ensuit que :

$$Y_K = Y_0 K_K = Y_0 e^{rK}$$

Cette équation est la relation de base qui établit le lien théoriquement entre revenu et niveau éducatif. Le taux d'actualisation  $r$  est également le taux de rendement de l'investissement éducatif. La forme logarithmique permet d'établir une relation linéaire entre niveau d'étude  $s$  et le logarithme du revenu :

$$\ln Y_K = \ln Y_0 + rK$$

Dans ce cadre d'analyse, le coefficient  $r$  peut être interprété comme le rendement de l'investissement éducatif. En effet,  $r$  mesure en terme relatif ce que rapporte une année d'étude supplémentaire.

Mais la simplicité de cette relation n'est pas gratuite. Celle-ci n'est valide que sous une certaine hypothèse [2]. D'abord on suppose que l'éducation n'engendre aucun coût direct. Même dans les pays où l'accès à l'école et à l'université est gratuit tel que le Maroc, l'élève comme l'étudiant sont en général dans la nécessité de supporter certains coûts afin d'arriver à poursuivre dans des conditions normales leur scolarité. De plus, dans certains pays comme les États-Unis, l'enseignement supérieur est très onéreux. Pour sortir de l'impasse, Mincer [2]

emprunte l'hypothèse postulé par Hanoch (1967) qui stipule que l'individu gagne de l'emploi à temps partiel ce qui compense le total des frais éducatifs.

Ensuite on suppose que le log revenu évolue linéairement avec l'investissement éducatif. On verra plus tard que les données empiriques sont loin de valider une telle hypothèse.

Ensuite, on fixe la durée de vie professionnelle. Ainsi, le rallongement de la durée de formation ne diminue pas le nombre de flux financiers à recevoir mais cause la baisse de leur valeur actuelle. Également, on supposera que la dépréciation du capital humain durant les années d'étude est nulle.

Pour terminer, le modèle scolaire n'est valide que si on suppose que l'investissement en capital humain prend fin une fois la formation initiale terminée. Autrement dit l'étudiant n'investit pas durant sa vie professionnelle, n'apprend pas par l'expérience et ne s'inscrit pas dans des programmes de formation continue. Cela permet de poser le revenu comme constant le long de la vie professionnelle. On néglige par la même occasion l'impact des changements économiques.

Le même résultat obtenu par Mincer a été démontré par Chiswick et Becker (1966) [3]. Après l'exposition de leur raisonnement ils parviennent à exprimer le log du revenu en fonction de la durée de l'investissement.

$$\log E_i = a + \bar{r}K_i v_i$$

Où  $\log E_i$  est le logarithme du revenu  $\bar{r}$  le rendement moyen et  $K_i$  la durée de l'investissement éducatif.

L'application de ce modèle aux données relatives aux revenus des hommes blancs en 1959 confirme l'effet positif de l'investissement éducatif sur le revenu [3]. Deux régressions séparées ont été estimées afin de distinguer la région sud de la région non-sud. Les résultats montrent que dans le sud le coefficient de régression est compris entre 7% et 9% alors qu'en région non-sud il est compris entre 5% et 8%. Les taux varient légèrement quand on passe d'un niveau éducatif à un autre. Le  $R^2$  est assez faible, compris entre 7% et 16%.

Les estimations tirées des données sur les revenus de 1959 des hommes blancs, âgés entre 15 et 64 ans, (les fermiers et les étudiants non inclus) présentent des résultats aussi faibles [4], le coefficient  $R^2$  se limite à 7% et le coefficient de régression (pente de régression) est faible, proche de la moitié des taux de rendement internes estimés suivant d'autres méthodes de calcul plus directs.

Le pouvoir explicatif du modèle reste donc faible. Les hypothèses très fortes posées au départ sont probablement responsables de la performance peu satisfaisante du modèle classique de rendement scolaire. En effet, le modèle scolaire peut être considéré comme un cas particulier d'investissement en capital humain. L'apprentissage par expérience complète l'investissement éducatif et influence par conséquent le revenu observé.

Effectivement, au cours des premières années de travail, la personne relativement débutante et récemment diplômée continue à investir dans son capital humain à travers l'apprentissage par expérience. Il s'ensuit que le premier salaire perçu correspond en fait au revenu potentiel exigé après  $s$  années d'études diminué du coût de l'apprentissage par l'expérience (coût de

l'investissement). Cela est d'autant plus vrai que la formation appliquée est à caractère générale et donc bénéfique à toutes les entreprises.

Supposons que la productivité potentielle d'un nouvel employé récemment diplômé est de  $Mp_K$ . Sur un marché de concurrence pure et parfaite en équilibre l'employé est rémunéré à hauteur de sa productivité marginale tel que  $Mp_K = Y_K = E_K$ . Durant la période d'apprentissage, la productivité réelle de l'employé notée  $Mp_K^r$  est inférieure à sa productivité potentielle  $Mp_K$  en raison du temps perdu. Le processus d'apprentissage occasionne également des pertes et des casses  $T$ . Selon Becker [1] une formation à caractère général n'est pas bénéfique à la firme sur un marché de travail concurrentiel. L'employé doit donc supporter à lui seul le coût de l'investissement. Durant les premières années de sa carrière, il reçoit une rémunération à hauteur de sa productivité réelle diminuée des pertes et des casses occasionnées. Ainsi le salaire qu'on observe ne reflète pas son capital humain accumulé durant les années d'étude.

$$Y_0 = Mp_K^r - T = Mp_K - (Mp_K - Mp_K^r) - T = Mp_K - I < Mp_K$$

Où  $I = (Mp_K - Mp_K^r) + T$  est le coût total de l'investissement.

Les revenus de début de carrière sous-estiment le rendement de la formation salariale. Le caractère inobservable du salaire potentiel compromet les résultats de l'étude empirique. A mesure que les années passent, le salaire perçu à chaque période sera augmenté du rendement des investissements déjà effectués moins le coût de l'investissement consenti. Il devient par conséquent crucial de prendre en compte l'effet de l'expérience afin de corriger les insuffisances et éliminer certains biais inhérents à l'ancien modèle.

### **2.1.2 Modèle de rendement de l'investissement scolaire et post scolaire : modèle développé par Mincer [2].**

Après les études, le travailleur entame sa vie professionnelle à la  $j^{ième}$  année. Il cherche dans ses débuts à améliorer ses compétences et peut-être à accumuler plus d'informations en relation avec son travail, ce qui lui coûte en termes d'argent et/ou en termes de temps. On note  $I_j$  le coût de l'investissement post scolaire. Le revenu observé  $Y_j$  en  $j$  est un revenu net des coûts d'investissement. Ainsi le revenu potentiel  $E_j$  n'est pas observé. Durant la première année d'expérience où  $j = 0$  on s'accordera à dire que le revenu perçu est donné par :

$Y_0 = Y_K - I_0$ , où  $Y_K = E_K$  est la capacité de gain initiale après  $s$  années d'études.

Le revenu observé l'année suivante dépendra également du coût d'investissement en capital humain dans le cas où le travailleur fait le choix de continuer à investir mais aussi du rendement de l'investissement passé. Ainsi le revenu observé en  $j = 1$  est donné par l'équation suivante :

$$Y_1 = Y_K + rC_0 - C_1$$

et de manière générale

$$Y_j = Y_K + \sum_{t=0}^{j-1} r_t I_t - I_j = E_j - I_j$$

Nous appellerons  $Y_K$  le capital humain scolaire car il rémunère le capital humain accumulé durant les années de scolarité. Le revenu net qui est au départ inférieur au capital humain scolaire, augmente tant que l'investissement net est positif  $I_j$  et qu'il baisse d'une année à l'autre  $I_{j+1} - I_j < 0$  en effet :

$$\Delta Y_j = Y_{j+1} - Y_j = r_j I_j - (I_{j+1} - I_j) = r_j I_j + I_j - I_{j+1} > 0.$$

Maintenant si l'investissement augmente d'une année à l'autre tel que  $I_{j+1} - I_j > 0$  alors le revenu net augmente aussi, si le taux de rendement  $r_j$  de l'investissement est supérieur au taux de variation de l'investissement [2]. En effet :

$$\Delta Y_j = r_j I_j - (I_{j+1} - I_j) = r_j I_j - \frac{(I_{j+1} - I_j)}{I_j} I_j = \left( r_j - \frac{(I_{j+1} - I_j)}{I_j} \right) I_j$$

Par conséquent :  $\Delta Y > 0$  si  $r_j > \frac{(I_{j+1} - I_j)}{I_j}$

Cependant, l'hypothèse d'un investissement décroissant est plus plausible. En effet, Becker [5] explique que le temps consacré aux études baisse à mesure qu'on avance dans l'âge et ceci pour deux raisons. Premièrement, la retraite ou encore la mort font que la durée de récupération du rendement sur investissement est limitée. Celle-ci rétrécit donc avec le temps qui passe. Il devient par conséquent moins attrayant d'investir à un âge avancé. Deuxièmement, avec l'accroissement du capital humain, le manque à gagner augmente et il devient alors de plus en plus coûteux d'investir.

Le revenu net continue à augmenter à un rythme croissant ou décroissant (selon le cas) jusqu'à ce qu'il atteigne le maximum au point où l'investissement net est nul. Lors de son accroissement, le revenu net après quelques années d'expérience notées  $\hat{j}$  dépasse le revenu potentiel initial  $Y_K$  perçu après  $K$  années d'étude par une personne qui décide de ne pas continuer à investir. L'idée de Mincer [2] est que si on arrive à déterminer le nombre d'années  $\hat{j}$  on saurait lire directement à partir du profil de revenu net observable la valeur du revenu potentiel initial après  $K$  années d'étude  $Y_K$ . Toute la question est de savoir comment estimer le nombre d'années nécessaires  $\hat{j}$  pour que le revenu net perçu soit égal au revenu potentiel initial  $Y_K$  ? En fait, la limite supérieure de  $\hat{j}$  peut-être estimée à partir de l'équation suivante :

$$Y_j = Y_K + r \sum_{t=0}^{j-1} I_t - I_j = Y_K$$

Donc  $r \sum_{t=0}^{j-1} I_t = I_j$

Ici on suppose que  $r_j = r$  ne varie pas d'une année à l'autre.

On peut faciliter les calculs en supposant que l'individu investit le même montant durant chaque période depuis  $t = 0$  et jusqu'à  $t = \hat{j}$ . En se faisant l'équation se transforme comme suit :

$$Y_K + r \sum_{t=0}^{j-1} I_t - I_j = Y_K \Rightarrow Y_K + rI_j\hat{j} - I_j = Y_K \Rightarrow rI_j\hat{j} = I_j \Rightarrow \hat{j} = \frac{1}{r}$$

Ainsi, quand le taux d'intérêt est supérieur à 10% , le revenu courant d'une personne ayant investie durant les années d'expérience rattrape en moins de 10 ans le revenu courant d'une personne du même niveau d'étude mais n'ayant jamais entrepris un investissement postsecondaire en capital humain. Ainsi, pour un taux d'intérêt de 10% et après une durée de 10 ans, les revenus courants des deux individus deviennent comparables.

Supposer le coût d'investissement constant est discutable. Cela contredit les résultats théoriques. Il s'agit d'une hypothèse nécessaire à la simplification des calculs.

Les bases de données contiennent des informations sur des revenus qui rémunèrent des parcours scolaires différents et des années d'expériences inégales. Le modèle scolaire simplifié n'est pas directement applicable compte tenu de la nature des données, autrement il conduirait à des estimations erronées. De plus, le revenu potentiel après s années d'études n'est pas directement observable. Une première idée serait de regrouper les données en fonction de l'expérience afin de neutraliser l'effet de l'investissement postsecondaire. Pour un taux de rendement de  $r = 0.1$ , le revenu potentiel rémunérant uniquement les années d'étude est observable après dix ans d'expérience. Ainsi, les personnes ayant accumulées dix ans d'expérience sont regroupées. L'estimation du modèle simplifié à partir des données utilisées par Mincer [4] conduit enfin à des résultats plus proches des estimations obtenues par des méthodes plus directes.

La validité de l'interprétation des estimations obtenues est tributaire à la pertinence du choix du taux de rendement. Un modèle plus élaboré qui tient compte de l'expérience sera plus adapté à la nature des données disponibles et nous évitera le caractère arbitraire inhérent au regroupement de données.

On pose  $K_j^{ex} = \frac{I_j}{E_j}$ . Le ratio  $K_j^{ex}$  peut être interprété comme une fraction du temps dévouée au développement des compétences avec  $I_j = E_j K_j^{ex}$ . Ce ratio vient donc en diminution du revenu courant net avec  $Y_j = E_j(1 - K_j^{ex})$ . Partant du fait que :

$$E_j = E_{j-1} + rI_{j-1} = E_{j-1} (1 + rK_{j-1}^{ex})$$

Cela implique par récurrence que

$$E_j = E_0 \prod_{t=0}^{j-1} (1 + r_t \hat{K}_t^{ex})$$

Pour tout  $K_j^{ex} \leq 1$  et  $r$  relativement faible on peut écrire  $E_j$  sous forme logarithmique comme suit :

$$\ln E_j = \ln E_0 + \sum_{t=0}^{j-1} r_t K_t^{ex}$$

Ce qui implique :

$$\ln Y_j = \ln E_0 + \sum_{t=0}^{j-1} r_t K_t^{ex} + \ln(1 - K_j^{ex})$$

Si on suppose maintenant que durant les années d'études  $K_j^{ex} = 1$ , on se retrouve avec une expression qui représente une extension du modèle scolaire :

$$\ln Y_j = \ln E_0 + r_K K + r_p \sum_{t=0}^{j-1} K_t^{ex} + \ln(1 - K_j^{ex}).$$

On pose  $\widehat{K}_t^{ex} = \sum_{t=0}^{j-1} K_t^{ex}$  le temps cumulé durant les années d'expérience. Le modèle devient :

$$\ln Y_j = \ln E_0 + r_K K + r_p \widehat{K}_t^{ex} = \ln Y_K + r_p \widehat{K}_t^{ex}$$

L'expression économétrique finale du modèle de Mincer est donnée par :

$$\ln Y_i = a + r_K K_i + b EX_i + c EX_i^2 + \epsilon.$$

Où  $K$  est le nombre d'années d'étude et  $EX$  est le nombre d'années d'expérience professionnelle passées sur le marché du travail. Le coefficient de régression  $b$  mesure le taux de rendement puisqu'il est défini par  $r_K = \frac{\partial \ln Y}{\partial K}$ . Remarquons que le modèle prend une forme quadratique.

### 3 Les tests empiriques de la relation salaire-revenu

Comment réagit le modèle de gain de Mincer face aux données empiriques ? Les études empiriques dont l'objet est d'estimer le rendement scolaire suivant la méthode de Mincer sont très nombreuses. Elles constituent une littérature qui a connu une évolution considérable.

Dans plusieurs études, le modèle classique de gain de Mincer ou une version modifiée de ce modèle sont appliqués aux données relatives aux revenus d'un échantillon de la population active [6] [7] [8] [9] [10] [11] [12] [13]. Certaines études portent sur un seul pays tandis que d'autres présentent une analyse comparative entre différents pays. Dans la version modifiée du modèle on retrouve en général les éléments du modèle classique ainsi que d'autres variables de contrôle tel que le genre, la zone de résidence, l'appartenance ethnique, la couleur de la peau et l'appartenance sociale. L'application de ces modèles aux données de revenus a montré un effet positif de l'éducation sur le revenu en Uruguay [6], au Brésil [7], au Venezuela [8], en Colombie [9], en Turquie [13] au Mexique [10], au Pakistan [11], en Chine [12] ainsi qu'un nombre considérable de pays.

Ces exemples sont loin d'être représentatifs de la littérature très vaste qui traite le sujet. Le coefficient de régression varie considérablement d'un pays à l'autre. Il est estimé à 9.2% en Uruguay [6], à 11.4% au Venezuela [8], à 15.65%, au Mexique [10], à 5.1% au Pakistan [11], à 8.8% en Chine [12] à 16.32% au Brésil [7] et à 8.8% en Turquie [13]. Néanmoins, ces taux ne sont pas comparables. En effet, les échantillons ne sont pas toujours représentatifs à l'échelle nationale. Par exemple, l'étude menée en Uruguay ne concerne que la zone urbaine.

Le coefficient estimé pour le Venezuela est vrai pour une population masculine. De plus, les modèles appliqués ne collent pas nécessairement avec le modèle classique et les modifications appliquées ne sont pas toujours les mêmes.

Il serait intéressant de profiter de l'immense littérature qui existe à ce sujet afin de mener une analyse comparative des taux de rendement éducatif des différents pays du monde et obtenir à la fin une vision plus globale de l'éducation dans le monde. Dans cette optique, des rapports de synthèse ont été publiés par la banque mondiale. Le rapport publié en 2018 [14] montre que le taux de rendement d'une année d'étude supplémentaire est de manière générale de 8.8%. Ce taux est à la base de 705 estimations menées sur la période 1950-2014 [14]. Le taux de rendement des pays à faible revenu est de 9.3% contre 8.2% pour les pays à revenu élevé. Le taux de rendement éducatif des pays de l'Amérique latine est le plus élevé (11%), vient en deuxième position les pays subsahariens (10.5%) suivis des pays de l'Asie de l'Est et du pacifique (8.7%). Les économies avancées sont en quatrième position (8%) suivies de l'Europe et de l'Asie centrale (7.3%). Les résultats des estimations sont à considérer néanmoins avec précaution puisqu'il s'agit d'une collecte de données à l'échelle internationale mobilisant plusieurs sources certaines plus fiables que d'autres. Cependant, ces comparaisons sont construites sur la base des estimations qui ne sont pas parfaitement comparables. En effet, les modèles de gains appliqués ne sont pas identiques. Ainsi, les différences des taux peuvent être dues aux différences dans les méthodes utilisées.

Les travaux de synthèse ou d'analyse comparative qui s'appuient sur les résultats des différentes études sélectionnées à partir d'une large revue de littérature présentent certaines difficultés liées à la comparabilité des données. En générale, les chercheurs qui étudient le sujet utilisent des méthodologies différentes. Dans plusieurs cas on remarque l'application des versions modifiées du modèle classique de Mincer. De plus, les différentes bases de données utilisées ne sont pas forcément représentatives. Pour des raisons financières on privilégie parfois la collecte de données facilement accessibles en se limitant par exemple à la zone urbaine.

### **3.1 Les rendements des investissements éducatifs dans le monde : une analyse comparative**

Les taux de rendement de 139 pays autour du monde publiés dans le document de travail de la banque mondiale de 2014 sont comparables [15]. Les estimations sont obtenues par l'application d'un modèle de gain unique à une large base de données (International income distribution data) constituée à partir d'une compilation des données provenant de 819 enquêtes. Les résultats de ce rapport peuvent être préférés à d'autres rapports plus récents pour une question de comparabilité.

Le modèle classique de Mincer a été utilisé. Les résultats montrent que les systèmes éducatifs les plus rentables appartiennent au Rwanda, en Afrique du Sud, en Namibie, en Chine et au Burundi qui affichent respectivement les taux de 22.4%, 21.1%, 18.5%, 18.3%, et 17.3% [13]. En bas du classement on trouve l'Afghanistan, l'Arménie, la Fédération Russie, la Guyane et l'Iraq qui affichent respectivement les taux les plus faibles de 1.6%, 2.2% 2.6% 3.3% et 3.4%. Les femmes du Rwanda sont celles qui bénéficient le plus de leur éducation. La rentabilité de leur investissement éducatif est de 24.4%. Celle-ci correspond au meilleur taux de rendement des 139 pays concernés par l'étude. Les femmes de l'Afghanistan connaissent le taux de

rendement le plus faible qui atteint 3%. C'est encore au Rwanda qui abrite les hommes les plus avantagés avec un taux de rendement de 20.8%. En Arménie, on rencontre les hommes qui bénéficient le moins de leur éducation avec un taux de rendement de 0.8%.

Au niveau régional, les pays de l'Afrique subsaharienne connaissent le taux de rendement moyen le plus élevé qui est de 12.4%. Vient en deuxième position les économies à revenu élevé avec un taux moyen de 10% suivies des pays de l'Asie de l'Est et pacifique avec un taux de 9.4%. En dernière position on retrouve le Middle-East et les pays de l'Afrique du Nord avec un taux de rendement moyen de 7.3%.

A l'échelle de tous les pays concernés par l'étude, le taux de rendement moyen pour l'ensemble de l'échantillon est de 10.1%. Les femmes connaissent un rendement plus élevé de 11.7% contre 9.6% pour les hommes. Entre 1980 et 2013 le taux de rendement moyen connaît une tendance plutôt baissière. Il passe de 13.3% entre 1980 et 1985 à 10% entre 2011 et 2013. Il atteint son niveau le plus bas de 9.6% en 2006-2010. Toutefois, il faut rester très prudent dans l'usage et l'interprétation de tels résultats. En effet, la distribution du nombre d'enquêtes utilisées sur les différentes périodes est très déséquilibrée. On constate que le nombre d'enquêtes utilisées augmente avec les périodes. Ainsi dans la période 1980-1985 le nombre d'enquêtes ne dépasse pas 12 contre 238 dans la période 2006-2010.

### **3.2 L'éducation est plus rentable pour les femmes.**

Afin d'estimer un taux de rendement spécifique pour les femmes et un taux de rendement spécifique pour les hommes on doit procéder en deux étapes. La première étape consiste à séparer l'échantillon en deux sous-échantillons afin de distinguer les hommes des femmes. Ensuite, dans la deuxième étape, le modèle classique de gain ou une version modifiée est appliquée séparément pour chacun des deux sous-échantillons. L'objectif est de comparer la rentabilité entre les deux genres.

De manière générale, on constate que l'éducation est plus rentable pour les femmes. Les documents de travail publiés par la banque mondiale en 1985 [16] puis en 2014 [15] et enfin en 2018 [14] montrent que le taux de rendement de l'investissement éducatif est plus élevé pour les femmes, comparé aux hommes. Dans le document de travail publié en 2018, [14] la différence est à environ 2 points de pourcentage. Dans le document de travail publié en 2014, sur 139 économies concernées, le taux de rendement en moyenne pour les femmes est de 11.7% contre 9.6% pour les hommes.

Ces résultats concordent avec certaines études qui ont pu montrer également que le taux de rendement de l'investissement éducatif est plus élevé pour les femmes en Turquie [13] (13.4% contre 8.3%) en Uruguay [6] (10.2% contre 9.1%), au Pakistan [11] (8.3% contre 4.5%) et au Maroc (14.5% contre 10.35%) [17].

Cette différence ne prouve en rien que les salaires des femmes sont plus élevés. C'est généralement le contraire qui ressort des études. En revanche cet avantage au profit des femmes montre que l'éducation est un bon investissement pour elles.

### **3.3 L'investissement éducatif est plus rentable dans le secteur privé comparé au secteur public.**

En Uruguay, l'analyse des données de 1989 montre que le secteur privé permet de mieux rentabiliser l'investissement éducatif [6]. Le taux de rendement de l'éducation est de 10.9% dans le secteur privé contre 6% dans le secteur public. A Caracas, située au Venezuela [8] le secteur privé enregistre un taux de rendement scolaire de 10.4% en 1975 contre 8.5% pour le secteur public. Cependant le taux de rendement dans les deux secteurs privé et public en 1984 se rapprochent ; ils affichent respectivement des taux de 9.8% et 9.7%. L'exploitation des données collectées en 1988 montre qu'à Bogota en Colombie, le secteur public comme le secteur privé connaissent des taux de rendements éducatif proches. Ils sont respectivement de l'ordre de 9.9% et 9.7% [9].

Ces résultats contredisent l'étude menée pour la Turquie [13] où le rendement éducatif dans le secteur public est plus élevé. Dans le cas de la Turquie, le secteur public possède la particularité de mieux rentabiliser l'investissement éducatif comparé au secteur privé ce qui n'est pas habituellement observé dans d'autres pays. Les salaires dans le secteur privé correspondent pour la majorité au salaire minimum alors que dans le secteur public les traitements dépassent le seuil du minimum. Cependant, ces résultats sont à relativiser puisque depuis 2011, le nombre important des réfugiés syriens a substantiellement ramené à la baisse les salaires du secteur privé. Il est à noter que le niveau des salaires du secteur privé ne promet pas de s'améliorer même après avoir retiré les réfugiés de la population étudiée si on inclut juste après les 30% jusqu'à 35% des salaires du secteur informel.

Le document de travail publié en 2018 par la banque mondiale [14] suggère qu'en moyenne l'investissement éducatif est mieux rentabilisé dans les secteurs privés comparés aux secteurs publics. La moyenne est estimée à partir d'un certain nombre d'études concernant différentes économies.

### **3.4 L'effet de la zone géographique**

En Chine, il a été enregistré à partir des données datant de 2010 que la zone urbaine est plus rentable comparée à la zone rurale [12]. En effet le taux de rendement de l'investissement en zone urbaine est de 12.2% tandis qu'il ne dépasse pas en moyenne 2.3% en zone rurale. Le rendement éducatif dans l'est est de 10.7% alors qu'il s'établit en moyenne à 5.4% dans la région ouest. En Finlande, l'exploitation des données de la « longitudinal census file » et la « longitudinal employment statistics file » montre que le rendement éducatif chez les salariés est de 12.4% dans la zone urbaine contre 10.3% dans la zone rurale tandis que chez les entrepreneurs le rendement de l'investissement éducatif est de 14.4% dans la zone urbaine contre 12% dans la zone rurale [18]. On conclut que les entrepreneurs et salariés de la zone urbaine rentabilisent mieux leur investissement éducatif comparé au entrepreneurs et salariés de la zone rurale. Au Pakistan, les estimations par région et par genre sont contrastées [11]. Le taux de rendement de l'investissement éducatif chez les hommes estimé à partir des données de l'enquête 2006-2007 est de 4.7% dans la zone urbaine contre 5.6% dans la zone rurale. Chez les femmes, c'est l'inverse qui est constaté, le taux de rendement de l'investissement éducatif dans la zone urbaine est de 19.5% contre un rendement presque nul dans la zone rurale.

Au Brésil, les taux de rendement diffèrent largement d'une région à une autre [19]. Par exemple le taux de rendement de l'investissement éducatif est de 6.56% à Sainte-Catherine contre 10.17% à Piauí. On conclut alors que les rendements éducatifs diffèrent d'une région à une autre.

### **3.5 Que signifie des taux de rendement différents ?**

Commençons avec ce que ne veut pas dire des taux de rendement différents. Le taux de rendement plus élevé pour les femmes par rapport aux hommes ne veut pas dire que les femmes sont mieux payées que les hommes. De même, un taux de rendement de l'investissement éducatif plus élevé dans un pays ne reflète en aucun cas le niveau des revenus dans ce pays. D'ailleurs nous avons constaté que les pays à revenus élevés ne sont pas les mieux classés par rapport au rendement [15]. En fait, un taux de rendement plus élevé reflète une inégalité plus importante de situation entre les personnes sans éducation ou à niveau d'étude faible et les personnes plus instruites.

Aussi, la comparaison entre différents taux de rendements ne permet pas de distinguer correctement entre les valeurs des différents systèmes éducatifs. En effet, deux marchés différents peuvent valoriser différemment un seul et même système éducatif. Ainsi, une meilleure comparaison peut être obtenue en estimant le rendement pour des personnes ayant complétées leur éducation dans différentes régions ou pays mais qui travaillent dans le même marché.

Les données concernant des migrants provenant de différentes régions du Mexique et travaillant à Sao Paulo ont été exploitées afin de mener une analyse comparative entre les systèmes éducatifs des différentes régions du Mexique [19]. Les migrants concernés par l'étude ont tous terminé leurs études dans leur région d'origine. Les résultats obtenus après l'application d'une version modifiée du modèle classique de Mincer, montrent une différence significative entre les taux de rendement des différentes régions et les taux de rendement éducatif pour les migrants qui travaillent à Sao Paulo. Par exemple, le taux de rendement éducatif à Marancho est de 10.16% tandis que le rendement éducatif des les migrants qui travaillent à Sao Paulo et qui ont terminé leurs études à Marancho ne dépasse pas 3.97%. Par rapport au taux de rendement par région, Espirito paraît moins bien classé que Maranchao avec un taux moyen ne dépassant pas 8.49%. Le classement d'Espirito s'améliore nettement si on considère le taux de rendement éducatif des migrants qui atteint 11.23%. Cependant, rien ne garantit que l'échantillon des migrants est représentatif de la population d'origine. Celui-ci peut présenter un biais de sélection. Après l'application de la solution classique de Heckman utilisée pour la correction du biais de sélection, les résultats obtenus montrent toujours un écart considérable entre les taux de rendement des migrants et les taux de rendement des différentes régions. Ainsi, le classement entre pays ou région change suivant la méthode de calcul. Le taux de rendement des migrants provenant des différentes régions sert à construire un indice de la qualité de l'enseignement assurée par les différents systèmes éducatifs.

Afin de mieux comparer les systèmes éducatifs des différents pays arabes on privilégiera l'usage des données concernant des migrants arabes ayant terminés leurs études dans leur pays d'origine et qui travaillent tous dans le même marché de travail. Des données contenant des informations sur des migrants arabes qui travaillent aux Etats-Unis et qui ont terminé leurs études dans leur pays d'origine ont été exploitées [20]. Les données utilisées concernent

deux coupes datant de 1990 et 2000. Une version modifiée du modèle de gain de Mincer a été appliquée. Les variables de contrôle supplémentaires introduites au modèle classique sont le nombre d'heures de travail, la compétence en anglais, le statut matrimonial, la santé (health limiting work), le lieu de résidence, ainsi que deux autres variables pour différencier les cohortes et les coupes. Le rendement éducatif des migrants arabes ayant terminés leurs études dans leur pays d'origine tout pays confondu est de 5.61% pour la coupe 1990 et 6.48% pour la coupe 2000. Le rendement éducatif par pays met à la tête du classement pour la coupe 1990 l'Algérie avec un taux de 13.75% suivie de la Palestine avec un taux de 8.07, vient ensuite l'Égypte avec un taux de 7.07%. En dernière position on trouve la Tunisie avec un taux de 1.85%. Le Maroc vient en septième position avec un taux de 5.11%. L'étude montre aussi qu'en partie ces différences de taux de rendement s'expliquent par la qualité de l'enseignement du pays d'origine. Ainsi, le Maroc est classé selon les données 1990 en septième position sur 12 pays. Selon les données 2000, l'Algérie préserve sa position à la tête du classement avec un taux de 9.09% suivie du Kuwait, vient par la suite le Maroc et le Liban en troisième position. Il s'agit d'une amélioration nette dans le classement pour le Maroc.

### **3.6 Le rendement éducatif au Maroc**

Un certain nombre d'études ont été menées dans le but d'estimer le rendement de l'investissement éducatif au Maroc. On remarque l'usage plus fréquent d'une version modifiée du modèle classique de Mincer. Une version modifiée du modèle reprend le modèle classique auquel on introduit des variables de contrôle telles que le genre [21], l'état matrimonial [21], la région de résidence [21] [17] la branche d'étude [17], la catégorie socioprofessionnelle [17], le secteur [17], la taille de l'entreprise [17], le type de contrat [17], le l'inverse du ratio de Mills [17] et les variables institutionnelles [22].

A partir des données de l'enquête nationale de l'emploi (l'ENE 2011), le taux de rendement au Maroc est estimé à 18.2% [21]. Dans la région de Casablanca, le rendement de l'investissement éducatif est estimé à 11.52%. Ce résultat est obtenu à partir des données de l'enquête conduite auprès des salariés durant le dernier trimestre 2020 dans la région de Casablanca [23]. Le taux de rendement de l'enseignement supérieur au Maroc est de 13.7%. Le résultat est obtenu à partir des données de l'enquête menée en 2020 auprès des salariées du secteurs privés diplômés des établissements publics [22]. L'introduction des variables institutionnelles dans le but de modéliser l'effet de l'établissement conduit à une baisse du taux de rendement qui atteint 11.79%.

Au Maroc comme dans plusieurs autres pays, l'investissement éducatif est plus rentable pour les femmes. En effet, à partir des données de l'enquête nationale réalisée en 2018 au Maroc, le taux de rendement est estimé à 14.5% pour les femmes contre 10.35% pour les hommes [17]. Selon des données datant de 2020, le taux de rendement à Casablanca, pour les femmes est de 12.4% contre 10.58% pour les hommes [20]. Cependant, un taux de rendement plus élevé n'est pas synonyme de salaires plus élevés. En fait, les femmes gagnent 44.8% moins que les hommes [21]. Parmi les diplômés du supérieur, les hommes gagnent 22.98% de plus que les femmes [21].

Les salaires diffèrent selon les régions. Habiter au Grand-Casablanca permet de gagner 2.6 fois ce que gagne son homologue qui habite d'autres régions [21]. Les Marocains de genre masculin résidant à Guelmim Oued Noun et à Ayoun Sakia Hamra sont les plus désavantagés

dans la mesure où ils gagnent respectivement 56.39% et 34.36% moins que les hommes habitant à la région orientale [17]. A Béni Mellal-Khénifra et à Tanger-Tétouan- Al Hoceima résident les femmes les moins avantagées recevant respectivement 46.35% et 39.91% de moins que les femmes résidantes dans la région orientale [17]. Toujours en comparaison avec les résidents de la région orientale, habiter à Rabat-salé-Kénitra apporte un gain supérieur de 6.1% pour les hommes, tandis qu'habiter au Grand Casablanca-Settat apporte un plus de gain de 12.91% pour les femmes [17]. Pour les diplômés du supérieur, habiter à Daraa-Tafilat, à Guelmim oud-noun et Laayoun Sakia Hamra apporte les plus grands désavantages salariaux puisque dans ces régions les diplômés du supérieur gagnent respectivement 18.25%, 14.52% et 14.25% de moins que leurs homologues habitant au grand-Casablanca-Settat [22].

Comparé aux diplômés des sciences sociales, l'obtention d'un diplôme en sciences d'ingénieur en technologie et sciences techniques, apporte un surplus de 14.82% pour les hommes et 9.8% pour les femmes [17]. Un diplôme en sciences de la santé est également avantageux pour les deux sexes avec une prime aux alentours de 10.5% de plus [17]. Cependant, pour les femmes, c'est le diplôme en science économique/ gestion/commerce qui apporte la plus grande prime avec 16.14% de plus que les diplômés des sciences humaines et sociales [17]. Ces résultats concordent avec les estimations réalisées à partir des données de l'enquête menée en 2020 auprès des salariées du secteur privé diplômés des établissements publics [22]. Ces estimations montrent que les diplômés en sciences humaines et sociales sont les plus désavantagés. Ils gagnent 29.46% de moins que les diplômés en science économique/gestion/ commerce.

Les hommes et les femmes travaillant dans le secteur privé non agricole gagnent respectivement 25.5% et 39.9% moins que les hommes et les femmes travaillant dans le secteur public ce qui présente un résultat surprenant [17]. Quant au secteur privé agricole, il est plus désavantageux pour les hommes sauf que la sanction baisse légèrement pour les femmes. Les hommes et les femmes travaillant avec un contrat à durée déterminée gagnent respectivement 14,44% et 24.64% de moins que les hommes et les femmes signataires d'un CDI [17].

#### **4 Limites du modèle de Mincer**

La littérature abondante sur le sujet manifeste la facilité de l'application de la méthode de Mincer. A partir de trois variables uniquement le revenu, la durée des études et l'expérience, l'estimation du rendement éducatif devient un jeu d'enfant. Mais cette simplification à l'extrême n'est-elle pas un avantage et un grand inconvénient en même temps. Nous avons présenté au niveau du fondement théorique les hypothèses sous lesquelles le modèle a été développé. Elles sont nombreuses et ne semblent pas toutes valides empiriquement. Comment alors s'assurer que les estimations générées par l'application de ce modèle ne sont pas aberrantes ? Dans le but d'améliorer la précision des estimations, une littérature critique s'est développée au sujet dudit modèle. Elle a mis en évidence un certain nombre d'erreurs et biais susceptibles d'affecter la qualité des estimations. Elle a proposé par la même occasion des solutions et corrections afin de dépasser ces limites. Dans cette deuxième partie nous présentons les différents défis relevés par l'analyse critique ainsi que les solutions suggérées afin de les dépasser.

#### 4.1 La qualité de l'éducation une variable manquante

Le rendement éducatif estimé à l'aide du modèle de Mincer est valable pour tous les établissements. Ainsi, quel que soit l'établissement d'origine, le rendement éducatif ne dépend que de la durée de formation. Il s'agit bien d'une insuffisance du modèle, puisque les établissements éducatifs sont dissemblables. Il serait absurde de penser que les diplômés des établissements prestigieux récupèrent le même rendement que les diplômés des établissements peu réputés du moment qu'ils ont accumulé le même nombre d'années d'études. Le rendement d'un même niveau d'études varie d'une école à une autre. Négliger le paramètre de la qualité de l'enseignement peut induire à des estimations biaisées [24]. La solution proposée par Berhman et Birdsall est d'introduire la variable manquante. La qualité de l'éducation devient alors un paramètre dans la fonction de gain classique. Berhman et Birdsall proposent une version modifiée du modèle de Mincer :

$$\ln Y_{KQ} = \ln Y_{0,0} + r(Q)K + bEX + cEX^2 + \epsilon$$

Où  $s$  est le nombre d'années d'études,  $r(Q)$  est le taux de rendement en fonction de la qualité,  $EX$  est le nombre d'années d'expérience et  $\epsilon$  est le terme stochastique. Le taux de rendement dépend alors de la qualité dispensée

Berhman et Birdsall reformulent le problème d'estimation du rendement comme suit

$$\ln Y_{KQ} = \ln Y_{0,0} + r^*K^*(K, Q) + bEX + cEX^2 + \epsilon$$

Où  $K^*(K, Q)$  peut prendre deux formes. Une forme quadratique  $K^*(K, Q) = r_0^* + r_1^*K + r_2^*Q + r_3^*K^2 + r_4^*Q^2 + r_5^*KQ$  ou une forme linéaire  $K^*(K, Q) = r_0^* + r_1^*K + r_2^*Q$ .

Le modèle développé admet un certain nombre d'hypothèses. En premier lieu, on suppose que la qualité de l'enseignement varie d'une région à une autre. Deuxièmement, la qualité n'incite pas les individus à changer de région de domicile. La troisième supposition stipule que la qualité dépend de l'allocation des ressources publiques spécifiques au secteur de l'enseignement à partir des recettes globales de manière à ce qu'il n'y ait aucun lien entre qualité dans une région et la charge fiscale supportée par un ménage de cette même région.

Les paramètres du modèle ont été estimés à partir des données tirées d'un sous-échantillon aléatoire de 6171 hommes, des ménages brésiliens recensés en 1970 [24]. La qualité de l'enseignement a été mesurée par le niveau moyen des enseignants (nombre d'années d'études des enseignants). Statistiquement la forme linéaire de  $K^*(K, Q)$  est plus performante mais les auteurs préfèrent la forme quadratique pour ses implications empiriques.

Les résultats des estimations montrent que la fonction de gains standard surestime amplement le rendement éducatif. Le rendement éducatif calculé à l'aide de la formule standard est de (20.5%) tandis qu'il ne dépasse pas 11.7% après avoir tenu compte de la qualité de l'enseignement (modèle quadratique). La validité des résultats dépend de la qualité de la spécification économétrique. En effet la relation entre taux de rendement et qualité est inconnue. Les deux formes linéaire et quadratique sont de suppositions uniquement. Il y a donc besoin d'un fondement théorique qui justifie la nature de la relation entre rendement éducatif et qualité de l'enseignement.

Une étude réalisée récemment au Maroc montre également que le rendement éducatif est surestimé quand on néglige l'effet de l'établissement [22]. A partir des données de l'enquête conduite en 2020 auprès des salariés du secteur privé diplômés des établissements publics, le taux de rendement de l'enseignement supérieur au Maroc est estimé à 13.7%. L'introduction des variables institutionnelles tel que le type d'accès (ouvert/ régulé), le taux d'encadrement, le corps administratif et la part des professeurs PES par établissement conduisent à une baisse du taux de rendement qui atteint 11.79%.

#### 4.2 La linéarité de la relation éducation-log revenu en question :

Le taux de rendement moyen de l'investissement éducatif estimé à l'aide du modèle de Mincer est valable pour tous les niveaux d'études. Ainsi, le rendement reste le même à chaque fois que la durée de formation est prolongée d'un an. La constance du rendement moyen est l'implication directe de la linéarité entre log revenu et niveau d'études. Dans quelles mesures les données empiriques valident-elles cette linéarité supposée ?

Afin d'éliminer toute restriction concernant la forme que prend la courbe log-revenu, des variables dichotomiques sont incluses pour chacune des dix-sept années de scolarisation. L'idée est de remplacer la variable  $s$  par dix-sept variables dichotomiques qui prennent la valeur de 1 si l'individu a passé ce niveau et 0 sinon. Cette méthode permet d'estimer avec précision pour chaque niveau d'étude un taux de rendement spécifique. Une méthode alternative consiste à introduire uniquement trois variables dichotomiques reflétant les différents cycles de l'enseignement. Le modèle de gain s'écrit alors comme suit :

$$\ln Y = a + bPRIM + cSEC + dSUP + eEX_i + fEX_i^2$$

$PRIM$  pour le niveau primaire,  $SEC$  pour le niveau secondaire et  $SUP$  pour le niveau supérieur.

Les différents taux spécifiques de rendement peuvent être déduits des paramètres  $(b, c, d)$  comme suit :

$$r(\text{primaire par rapport aux sans qualification}) = \frac{b}{K_p}$$

$$r(\text{secondaire par rapport au primaire}) = \frac{c - b}{K_s - K_p}$$

$$r(\text{supérieur par rapport au secondaire}) = \frac{d - c}{K_h - K_s}$$

A partir des données de la PNAD (1982) contenant des informations sur la force de travail brésilienne, Strauss et Thomas estiment le rendement de l'éducation brésilienne par niveau d'étude et pour trois macro-régions [25]. Les résultats obtenus rejettent l'hypothèse de linéarité de la fonction de gains de Mincer. La fonction du log-salaire conditionnel prend une forme plutôt convexe sans qu'elle soit pour autant lisse. Le rendement éducatif augmente avec le niveau d'étude. Par exemple, pour les hommes de la région Nord-Est, le rendement éducatif passe de 11% à 15% puis à 16% à mesure que le niveau d'études progresse du cycle primaire au cycle secondaire et du secondaire au post-secondaire. Sauf que cette convexité n'est pas

lisse ; le rendement éducatif des quatre premières années du cycle primaire pour les hommes du Sud atteint une moyenne de 12% essentiellement en raison du taux de rendement significativement élevé de la quatrième année (fin de la première partie du cycle primaire). Les personnes qui décident de quitter l'école l'année qui suit ou une année avant la fin du cycle primaire ne gagnent rien par rapport à la dernière année scolaire. Du reste, le taux de rendement de la deuxième étape du cycle primaire n'est que de 10%. Le rendement du postsecondaire au même titre que le secondaire est de 13% avec un taux de retour sur investissement très élevé pour la première et la quatrième année. La forme de la courbe du log-salaire conditionnel se caractérise donc par une convexité segmentée. Non seulement l'hypothèse du rendement salarial constant n'est pas vérifiée mais le rendement positif pour chaque année n'est plus la règle. Le salaire n'augmente pas forcément avec chaque nouvelle année d'études supplémentaire et dans certains cas le rendement enregistre un taux négatif. Ces derniers constats sont plutôt la propriété des premières années de chaque cycle (première année de la deuxième étape du primaire, première année du cycle secondaire...).

D'après Strauss et Thomas [25] trois raisons peuvent expliquer la convexité de la fonction de gain (log-salaire conditionnel). La qualité de l'enseignement a un rôle à jouer. S'il est vrai que le taux des écoles dispensant un enseignement de qualité est plus élevé dans les niveaux d'études supérieurs, les charges facturées en conséquence participeront dans le cas où l'accès au crédit est contraint au rationnement de certains enfants. Une deuxième cause plausible est la demande de travail qui peut jouer également un rôle important. Concernant ces deux premières raisons, l'inclusion des variables de contrôle tel que la qualité de l'établissement ou la région de résidence peuvent être utiles dans la modélisation du rendement. Cependant, la convexité de la fonction du log-salaire conditionnel peut refléter également les caractéristiques inobservables des aptitudes des personnes ayant fait le choix de poursuivre leurs études. Dans ce cas, les taux de rendement croissants ne représentent pas forcément la trajectoire potentielle d'un individu quelconque, relative aux choix de poursuite d'études. Les aptitudes des personnes qui quittent l'école dès la première année d'un cycle peuvent expliquer également le taux de rendement négatif observé pour certains niveaux éducatifs. En tout état de cause les deux auteurs s'accordent à déduire de la convexité du log-salaire conditionnel une augmentation des inégalités sociales et personnelles.

La non-linéarité de la relation log revenu-études paraît clairement dans un certain nombre d'études. En Thaïlande, Le taux de rendement varie d'un niveau éducatif à l'autre. Il est de 2.7% en primaire et passe à 5.215% au secondaire général puis il monte à 14.42% pour le bachelor [26]. Les résultats sont obtenus à l'aide d'une version modifiée du modèle classique de Mincer appliquée aux données de l'enquête nationale de la force de travail de 2010. Le rendement éducatif évolue considérablement du primaire au supérieur pour les migrants arabes installés aux Etats-Unis et ayant terminés leurs études dans leur pays d'origine [20]. En effet, le rendement éducatif des premières années d'études pour cette catégorie est estimé à 1.08% à partir des données de 1990 et à 1.15% à partir des données de 2000, tandis que le rendement de la treizième année d'étude et plus est estimé à 8.2% selon les données 1990 et à 10.26% à partir des données 2000. A partir des données de l'enquête PNAD 1982, l'application d'une régression plus flexible réservant à chaque niveau éducatif une variable dichotomique spécifique montre que le rendement n'est pas constant d'une année à l'autre au Brésil [7]. L'hypothèse de linéarité n'est pas très solide. Par exemple, le rendement de la

quatrième année est de 35.85% tandis qu'il affiche un taux de 17.99% pour la huitième année d'études.

Selon le document de travail publié par la banque mondiale en 2014, les études supérieures sont plus rentables [13]. Elles offrent une rentabilité moyenne de 15.2% contre 10.6% pour le primaire et 7.2% pour le secondaire. Dans la région de l'Afrique subsaharienne, on retrouve les taux de rendement les plus élevés pour les trois niveaux éducatifs. Dans les différentes régions du monde on constate la non-linéarité de la relation revenu-éducation. De manière générale, le rendement éducatif baisse quand on passe du primaire au secondaire puis augmente pour dépasser son niveau de départ dans le supérieur. Trois régions échappent à la règle. Les pays à revenu élevé ne connaissent pas de tels fluctuations. Le rendement ne fait qu'augmenter à mesure qu'on avance d'un niveau à l'autre.

### **4.3 Le biais d'endogénéité**

Comment interpréter le taux de rendement assez élevé selon certaines études tandis que dans d'autres travaux la mobilité intergénérationnelle est faible. N'est-il pas contradictoire de voir réuni un système éducatif rentable avec une société qui assiste à une faible mobilité entre classes sociales ? Si les pauvres ne réussissent pas à accéder à des classes sociales plus prestigieuses à l'aide de l'éducation, qui bénéficie alors de cette rentabilité ? Est-ce que les méthodes appliquées surestiment le taux de rendement ? Le modèle classique de Mincer permet-il d'estimer avec précision le rendement éducatif ?

Le modèle classique développé par Mincer néglige certaines variables susceptibles d'influencer le revenu individuel tel que les aptitudes. En effet, certaines personnes sont plus intelligentes que d'autres ou plus aptes dans certains domaines. L'omission d'une variable explicative pose problème quand la variable omise est corrélée avec une variable indépendante qui apparaît dans le modèle. La variable omise est alors capturée par le terme d'erreur qui devient de ce fait corrélé avec la variable indépendante. Il s'agit donc bien d'une transgression de l'une des hypothèses fondamentales de l'application correcte de la méthode des moindres carrés ordinaires. En effet, le modèle de régression par les MCO pose l'hypothèse de l'indépendance entre le terme d'erreur  $\varepsilon$  et le vecteur des variables explicatives  $X$  tel que  $corr(\varepsilon, X) = 0$ . La raison est simple. Si les personnes les plus intelligentes sont également les mieux formées, il y a alors une corrélation positive entre intelligence et éducation. Comment alors savoir si l'augmentation du revenu est due à l'intelligence ou à l'éducation. Il se peut que le coefficient associé à l'éducation ne reflète pas uniquement l'effet des années sur le revenu mais également une partie de l'effet de l'intelligence. L'omission du facteur aptitude peut entraîner alors une surestimation du rendement éducatif.

Les personnes plus habiles sont susceptibles d'investir plus dans leur éducation. Le modèle classique de Mincer peut présenter donc un problème d'endogénéité. Ce problème provient des caractéristiques non observables ayant un effet sur le revenu et qui sont corrélées avec l'éducation. Ces caractéristiques se retrouvent dans le terme d'erreur qui ne vérifie plus l'hypothèse d'indépendance nécessaire à l'application correcte de la MCO. Par exemple, le revenu peut être expliqué par les différences de niveau éducatif comme il peut être expliqué par les différences d'aptitudes. La différence de revenu observée entre deux niveaux d'études peut inclure également l'effet des aptitudes. La solution à ce problème est de trouver une méthode fiable pour mesurer la composante aptitude puis l'intégrer dans le vecteur des

variables explicatives. Le score des tests de qualification peut être considéré comme un proxy des aptitudes.

A partir des données de 1964 collectées auprès des vétérans des Etats-Unis de l'après deuxième guerre mondiale, le rendement éducatif est estimé à 5.02% [27]. Ce résultat est obtenu à l'aide d'une version modifiée du modèle classique de Mincer. Après l'introduction du score du test de qualification de la force armée dans l'équation de régression, le rendement éducatif baisse de 16.73% pour atteindre 4.18%. Ces estimations sont relatives aux années d'études accomplies avant le service militaire. Le score du test de qualification de la force militaire est un proxy de l'aptitude individuelle. Son incorporation dans la fonction de gain cause une baisse du taux de rendement. Ce résultat prouve dans une certaine mesure que le modèle classique de Mincer surestime le rendement. Cette conclusion est appuyée par les résultats d'une recherche menée sur les données de Bogota. Le taux de rendement de l'investissement éducatif à Bogota en 1988 estimé à 10.5% baisse jusqu'à 9.4% après introduction d'une variable mesurant la capacité de l'individu à raisonner [9]. Le taux baisse encore jusqu'à 8.5% après l'ajout de variable mesurant les performances cognitives

Notez que la baisse du rendement doit être interprétée avec prudence car celle-ci ne reflète pas uniquement l'effet net du biais d'endogénéité. L'incorporation de la variable aptitude ne résolve pas entièrement le problème car elle crée à son tour un nouveau problème d'endogénéité. En effet, l'aptitude individuelle peut-être corrélée avec la qualité de l'éducation [27]. Puisqu'il n'existe pas de variable qui capture l'aspect qualitatif de l'éducation, celle-ci se retrouve dans le terme d'erreur. Par conséquent l'hypothèse de l'indépendance entre l'erreur et la variable explicative est transgressée.

Selon Griliches et al, le niveau d'études accompli pendant et après le service militaire est peu corrélé avec les aptitudes [27]. Le coefficient de régression qui lui est associé est par conséquent un bon estimateur du rendement éducatif. Le taux de rendement de l'investissement éducatif réalisé durant et après le service militaire chez les vétérans des Etats-Unis est estimé selon les données 1964 à 5.28%. En introduisant la variable aptitude au modèle de régression le taux de rendement baisse de 10.73% seulement pour atteindre le niveau de 4.75%. On constate effectivement que la correction du biais d'endogénéité entraîne une baisse moins importante au taux de rendement. Cela appuie dans une certaine mesure l'hypothèse que l'investissement éducatif réalisé durant et après le service militaire est moins affecté par le biais d'endogénéité.

Le taux de rendement au Pakistan est estimé à l'aide du modèle classique à 5.1% [11]. En introduisant le score de Ravens test qui est une mesure d'aptitudes innées, le taux de rendement baisse à 4.8%. Aussi, le taux de rendement chez les hommes baisse de 4.5% à 4.3% et diminue également chez les femmes de 8.3% à 7.6%. Il permet donc de corriger les estimations du biais d'endogénéité. La baisse des taux de rendement est conforme aux prédictions théoriques mais le résultat n'est pas significatif [11]. La méthode directe de correction du biais d'endogénéité aboutit aux résultats attendus mais le résultat n'est pas significatif. Devons-nous conclure que les aptitudes affectent peu le revenu et que par conséquent le biais d'endogénéité ne pose pas de sérieux problème au modèle classique ? Autrement dit les erreurs dans les estimations données par la fonction de revenu de Mincer sont minimales et négligeables. N'est-il pas plus prudent de questionner dans un premier temps la pertinence des tests et indicateurs qu'on emploie en général pour mesurer les aptitudes

individuelles ? Se pose alors la question de la fiabilité des mesures d'aptitudes. Est-il possible de mesurer correctement les aptitudes individuelles. En fait les tests d'aptitudes ne sont pas efficaces pour mesurer les aptitudes innées reçues à la naissance et indépendantes de l'éducation. En général, les mesures d'aptitudes sont influencées par l'environnement et l'éducation [11].

Une solution alternative au problème d'endogénéité serait d'appliquer la méthode de la variable instrumentale. L'idée est de trouver une variable instrument qui vérifie deux conditions. La première condition exige la corrélation de la variable instrumentale avec l'éducation et les aptitudes. La deuxième condition exige la non-corrélation entre la variable instrumentale et le terme d'erreur [11]. Dans plusieurs cas on utilise l'origine sociale (caractéristiques de la famille d'origine) comme variable instrumentale.

Les caractéristiques de la famille d'origine sont des variables manquantes puisqu'elles ont un effet sur l'éducation. Si on suppose que les aptitudes sont transmissibles de génération en génération, il existe alors une corrélation entre l'éducation des parents et les aptitudes de la progéniture. On désigne par  $G_i$  les aptitudes d'une personne  $i$  et  $F_i$  les caractéristiques du père de la personne  $i$ . On peut alors écrire les aptitudes d'une personne en fonction des caractéristiques de ses parents tel que  $G_i = \pi F_i + G^n$  où  $\pi F_i$  représente les aptitudes héritées tandis que  $G^n$  représente la partie non héritée des aptitudes [7]. Ainsi, certaines caractéristiques de l'origine sociale peuvent remplacer les caractéristiques individuelles inobservables ayant un effet sur l'éducation et peut-être sur le revenu.

A partir des données de l'enquête PNAD 1982 le rendement éducatif au Brésil est estimé à 16.32% [7]. En rajoutant au modèle de base l'éducation du père le taux de rendement éducatif au Brésil baisse à 14.31%. Les taux de rendement de la quatrième et huitième année passent respectivement de 35.85% à 31.11% et de 17.99% à 15.4% [7]. Ainsi, une partie des différences de salaire récompense l'appartenance sociale ou bien les aptitudes héritées selon l'interprétation choisie.

Le rendement de l'investissement éducatif en Chine estimé à partir des données de l'enquête de 2010 est de 7.8% [12]. Après avoir introduit au modèle de régression l'éducation des parents ainsi qu'une variable dichotomique marquant ou pas la mort des parents à l'âge de 14 ans, le taux de rendement éducatif augmente jusqu'à 20.9%. Les deux nouveaux paramètres sont introduits comme variables instrumentales. Le taux de rendement chez les femmes monte également de 9% à 23.7%. De même pour les hommes, le taux de rendement augmente de 7.1% à 17.9%. L'estimation du rendement éducatif à partir des données de la Finlande montre également que l'application de la méthode de la variable instrumentale peut conduire à une augmentation des taux de rendements [18]. Sur la base des données de la longitudinal census file et de la longitudinal employment statistics file, les résultats de la méthode classique des MCO montrent que le rendement éducatif est de 10.4% pour les entrepreneurs contre 10.8% pour les salariés. Après la correction du biais d'endogénéité par l'introduction de l'éducation des parents au modèle, le rendement éducatif des salariés monte jusqu'à 13.2% tandis que le rendement éducatif récupéré par les entrepreneurs augmente à 16.2%.

On remarque que l'usage de la méthode de variable instrumentale gonfle les taux de rendement contrairement aux attentes. Ces résultats inattendus ne remettent pas forcément en cause le bien-fondé de la méthode mais attire notre attention sur la qualité de son application. En effet, le choix des variables instrumentales doit respecter les conditions mentionnées afin

de garantir un minimum de fiabilité. Dans la recherche réalisée en Chine les auteurs défendent leurs choix en démontrant que les variables choisies sont corrélées avec le niveau d'études [12], mais est-ce qu'ils le sont avec les aptitudes individuelles ? Dans la mesure où le choix d'une variable instrumentale n'est pas si évident, une troisième solution plus appropriée serait d'analyser les différences entre individus appartenant à la même famille [11]. Cette approche suppose implicitement que les personnes de la même famille partagent certaines caractéristiques. Par conséquent, l'effet des caractéristiques non observables partagées par les membres de la même famille est neutralisé.

L'origine sociale n'est pas uniquement une variable instrumentale ou une mesure des aptitudes léguées aux générations futures mais une variable manquante dans le modèle classique de Mincer qui peut être également source de transgression de l'hypothèse de l'indépendance du terme d'erreur. L'origine sociale capturée par le terme d'erreur puisqu'elle ne figure pas dans le vecteur des variables indépendantes du modèle classique, est en général corrélée avec le niveau d'études. Comment alors savoir si le revenu perçu par une personne est dû à son niveau d'études ou à l'appartenance à une catégorie sociale distinguée qui lui permet de faire jouer un réseau de relations important. Encore une fois l'omission de l'origine sociale peut conduire à une surestimation du rendement éducatif. La solution classique est de rajouter au vecteur des variables indépendantes, l'origine sociale.

L'introduction d'une variable capturant l'origine sociale dans le modèle classique baisse effectivement le taux de rendement des années d'études accomplies avant le service militaire par les vétérans des Etats-Unis de 24.5% [27]. Celui-ci passe de 5.02% à 3.79%. Le taux de rendement de l'investissement éducatif réalisé avant le service militaire baisse encore quand on prend en compte simultanément l'effet de l'aptitude et l'effet de l'origine sociale. La baisse enregistrée est de l'ordre de 34.66% [27]. L'incorporation du niveau éducatif du père ainsi que le niveau éducatif du père de l'épouse entraînent une baisse du taux de rendement de l'investissement éducatif au Brésil de respectivement 12.32% et 14.1% [7]. Quant au taux de rendement de la quinzième année, il baisse après l'ajout de ces deux paramètres de respectivement 11.2% et 17.64% [7]. On remarque avec surprise que le niveau éducatif du père de l'épouse a un effet plus important que le niveau éducatif du père. La relation entre les caractéristiques des parents de l'épouse et le revenu est justifiée par l'existence d'un appariement matrimonial tel que les revenus des mariés sont corrélés. Ceux qui touchent des revenus plus élevés se marient avec des femmes qui gagnent des revenus élevés. Puisque les femmes qui touchent les salaires élevés viennent des familles où le père est relativement plus éduqué, l'appariement fait sortir un effet indirect de l'éducation des parents de l'épouse sur le revenu.

D'après David Lam et Robert Schoemi [7], cette baisse du taux de rendement ne signifie pas forcément que le modèle classique de Mincer surestime le rendement éducatif. Cette baisse peut résulter également des erreurs de mesure de l'investissement éducatif. Les auteurs montrent qu'effectivement 80% de la baisse est due en fait à l'erreur. Ainsi on semble exagérer l'effet de l'origine sociale. Cela nous amène à se poser une question négligée jusqu'ici ; mesure-t-on correctement l'investissement éducatif ?

#### **4.4 L'erreur de mesure de l'investissement éducatif :**

Le niveau d'études capture mal le coût d'opportunité de l'investissement éducatif [28]. Il faut distinguer entre le niveau éducatif et la durée des études. En général, on suppose que le niveau éducatif reflète la durée des études ce qui n'est pas vrai dans tous les cas [28]. Le redoublement et l'abandon scolaire creusent un écart entre le niveau d'études atteint et le nombre d'années passées dans les études. Les redoublants passent plus d'années qu'il ne faut avant d'atteindre un niveau d'études donné. Les sortants du système scolaire avant l'obtention du diplôme quittent avant d'atteindre le niveau d'études suivant, ils sacrifient plus d'années d'études pour atteindre un niveau inférieur. Négliger les redoublements et les abandons conduit à sous-estimer le coût des études et par conséquent à surestimer le rendement éducatif.

Dans une étude menée à partir des données de l'enquête relative aux revenus en Indonésie, des chercheurs ont essayé d'estimer l'effet de l'abandon et du redoublement sur le rendement éducatif [28]. Afin d'estimer la durée des études approximative en prenant en compte l'abandon et le redoublement, les auteurs ont émis un certain nombre d'hypothèses. On suppose que toute personne qui a réussi à intégrer le niveau d'études suivant n'a connu aucun redoublement ou abandon dans les niveaux inférieurs, tandis qu'on applique à la personne qui n'a pas pu intégrer le niveau suivant un taux de redoublement et d'abandon approprié pour les niveaux inférieurs. Les résultats montrent que la méthode d'estimation classique de la durée des études surestime le rendement éducatif. Le taux de rendement du niveau primaire est surestimé de 38% à 78%, le rendement du niveau maternel est surestimé de 82% à 114%, le rendement du collège est surestimé de 14% à 24%. Ainsi la surestimation du taux de rendement peut être due à des erreurs de mesure de l'investissement éducatif.

#### **4.5 L'exploitation correcte des bases de données dans l'estimation du rendement éducatif.**

##### **4.5.1 Le biais de sélection**

Les estimations que nous avons présentées jusqu'ici sont basées sur les revenus de la force de travail active. L'échantillon est donc composé des individus qui ont un certain niveau éducatif ou sans qualification et qui participent sur le marché du travail. On exclut alors les personnes qui ont un certain niveau éducatif mais qui sont inactifs. L'échantillon est alors constitué suivant une règle de sélection.

Quand un échantillon est aléatoire, l'espérance du terme d'erreur d'une régression linéaire est nulle [29]. Mais si on procède à une sélection suivant une règle donnée rien ne garantit alors la nullité de l'espérance du terme d'erreur, celle-ci dépend alors de la règle de sélection. Si après application de la règle de sélection l'espérance de l'erreur est toujours nulle, on peut alors considérer le sous-échantillon sélectionné comme aléatoire. Il n'y a pas alors de biais de sélection. Dans le cas contraire on doit faire alors face au biais de sélection. L'exemple suivant est emprunté à Heckman [29] pour ses qualités pédagogiques.

Considérons les deux régressions linéaires suivantes :

$$\begin{cases} Y_i = X_i\beta + u_i \\ T_i = A_i\alpha + s_i \end{cases}$$

Où  $Y_i$  et  $T_i$  sont deux variables dépendantes expliquées respectivement par deux vecteurs de variables exogènes  $X_i$  et  $A_i$ . Les deux régressions ont pour terme d'erreur  $u_i$  et  $s_i$ . Dans le cas où l'échantillon  $I(i = 1 \dots N)$  est aléatoire on a  $E(u_i) = 0$  et  $E(Y_i|X_i) = X_i\beta$ . Supposons maintenant que tous les individus de l'échantillon ont un  $T_i$  qui est supérieur ou égale à 0. Par conséquent les valeurs  $Y$  des individus ayant un  $T$  inférieur à 0 sont manquantes. Il y a alors une sélection susceptible d'entraîner un biais. Dans ce cas [29]

$$E(Y_i|X_i, \text{règle de sélection}) = X_i\beta + E(u_i|T_i \geq 0) = X_i\beta + E(u_i|s_i \geq -A_i\alpha).$$

Sous certaine hypothèse et sur la base de résultats bien connus, Heckman montre que

$$E(u_i|s_i \geq -A_i\alpha) = \sigma_{YT}/\sigma_{TT}^{1/2}\lambda_i \text{ où } \lambda_i \text{ est l'inverse du ratio de Mills [29].}$$

Le modèle initial s'écrit alors comme suit

$$Y_i = E(Y_i|X_i, T_i \geq 0) + s'_i = X_i\beta + E(u_i|s_i \geq -A_i\alpha) + s'_i = X_i\beta + \frac{\sigma_{YT}}{\sigma_{TT}^{1/2}}\lambda_i + s'_i$$

Où  $s'_i$  est le nouveau terme d'erreur. Il vérifie  $E(s'_i|Y_i, \lambda_i, s_i \geq -A_i\alpha) = 0$ .

Corriger le biais de sélection revient à estimer pour chaque individu l'inverse du ratio de Mills, puis rajouter cette variable au modèle initial. Si les valeurs  $T_i$  sont disponibles aussi bien pour les individus ayant  $T_i \geq 0$  que pour les individus dont  $T_i < 0$  alors on peut estimer la probabilité d'avoir un  $T_i \geq 0$  en appliquant une analyse probit. Le résultat  $\alpha/\sigma_{TT}$  permet d'estimer le paramètre  $Z_i$  tel que  $Z_i = T_i\alpha/\sigma_{TT}$ . L'inverse du ratio de Mills est enfin estimé à partir du paramètre  $Z_i$  et intégré par la suite dans le modèle initial.

La procédure de correction du biais d'endogénéité de Heckman se compose de deux étapes essentielles [29]. Dans une première étape on utilise les données disponibles pour estimer un probit. Le résultat constitue la base de calcul de l'inverse du ratio de Mills. Puis on intègre les estimations obtenues dans le modèle initial afin d'obtenir des coefficients de régression corrigés du biais de sélection.

Dans le cas des bases de données utilisée dans l'estimation du rendement éducatif, la variable  $T_i$  est dichotomique. Elle prend la valeur de 1 si l'individu participe dans le marché de travail et 0 pour le cas inverse. La correction du biais de sélection passe d'abord par l'estimation de la probabilité qu'un individu participe au marché de travail en se basant sur certaines caractéristiques individuelles  $A_i$ . Ensuite on estime l'inverse du ratio de Mills qu'on rajoute à la fonction du gain.

Le taux de rendement de l'investissement en Chine estimé à partir des données de l'enquête de 2010 augmente après la correction du biais de sélection, il passe de 7.8% à 8.2% [12]. Le taux de rendement chez les femmes augmente de 9% à 9.7% tandis que chez les hommes il passe de 7.1% à 7.4% [12]. En Finlande le rendement éducatif après correction du biais de sélection suivant la méthode de Heckman baisse chez les salariés de 10.8% à 10.7% et chez les entrepreneurs de 10.4% à 10.3% [18]. Les résultats des deux recherches montrent que l'effet du biais de sélection est presque négligeable.

#### 4.5.2 Problèmes des revenus mixtes

Les revenus des entrepreneurs et des personnes qui exercent des professions libérales ne rémunèrent pas uniquement le capital humain. Une partie des bénéficiaires rémunère le capital physique. Comment dissocier entre les deux composantes ? Négliger les revenus non salariaux peut biaiser les estimations. En effet, certains employés créent leur propre emploi. Cette catégorie de travailleurs est concentrée chez le groupe de personnes faiblement diplômées [30]. On peut raisonnablement supposer après avoir tenu compte des différences d'entreprises et des aptitudes, que créer son propre emploi est plus avantageux pour les personnes à faible capital éducatif comparé au salariat [30]. Ainsi, le coût de l'investissement éducatif pour les personnes qui travaillent pour leur propre compte est relativement élevé. Ne pas y tenir compte peut induire à une surestimation du rendement éducatif. La solution est d'adapter le modèle de gain à la nature mixte des revenus. Le modèle défendu par Cheswick [30] est une solution raisonnable au problème.

On pose  $Y_i$  le revenu de la personne  $i$ . Ce revenu se compose d'un salaire reçu en travaillant pour un tiers et un revenu gagné en exerçant une activité pour son propre compte  $P_i^*$  tel que  $Y_i = w_i + P_i^*$ . On pose  $\alpha_i$  la proportion du revenu  $P_i^*$  rémunérant le travail et  $Y_{Li}$  la composante salaire du revenu total tel que  $Y_{Li} = w_i + \alpha_i P_i^*$ . Par conséquent le revenu d'une personne peut s'écrire sous forme  $Y_i = Y_{Li} + (1 - \alpha)P_i^*$ . Enfin on pose  $p_i = \frac{P_i^*}{Y_i}$  la proportion du revenu rémunérant le travail indépendant dans le revenu total. On peut montrer facilement avec un simple jeu de factorisation et de réarrangement que  $\log Y_i = \log Y_{Li} - \log[1 - (1 - \alpha)p_i]$ . On pose  $-\log[1 - (1 - \alpha)p_i] = Fp_i$  où  $F = E\left(\frac{-\log[1 - (1 - \alpha)p]}{p}\right)$ . Cheswick démontre que  $F$  peut être considéré comme constant pour  $0 < p_i < 1$ . Ainsi, la fonction log revenu peut s'écrire comme suit :

$$\log Y_i = \log Y_{Li} + Fp_i = a + r_K K_i + bEX_i + cEX_i^2 + Fp_i + \epsilon.$$

La fonction de gain classique est adaptée à l'estimation des rendements salariaux quand la base de données est composée par des salaires  $Y_{Li}$ . Cheswick [30] démontre que le modèle de gain peut être appliqué pour estimer le rendement salarial à partir d'une base de données à revenus mixtes après avoir rajouté le terme  $p_i$  à l'équation. On ne peut pas procéder plus directement sans passer par ce détour car  $\alpha$  n'est pas disponible.

A partir des données de l'enquête socio-économique menée en Thaïlande, Cheswick [30] estime le rendement salarial chez les employés hommes qui ne travaillent pas pour leur propre compte à 10.4%. Le rendement salarial baisse à 9.1% quand il prend en compte les revenus des personnes travaillant pour leur propre compte en appliquant l'équation de gain adaptée. Pour les femmes le taux de rendement passe de 14.5% (salié uniquement) à 13% (femmes travaillant pour leur propre compte y compris). On conclut que négliger les salaires des personnes travaillant pour leur propre compte conduit à une surestimation du rendement éducatif.

## 5 Conclusion

Le modèle Log-revenu-éducation de Mincer est une grande réussite dans l'économie du travail du XXème et XXIème siècle. Sa simplicité et son caractère pratique à l'usage

expliquent son utilisation très large et le développement d'une vaste littérature autour du sujet. A l'aide de la fonction de revenu de Mincer l'estimation du rendement éducatif est assez facile. Les estimations réalisées montrent qu'en général l'investissement en éducation est rentable ce qui justifie une relation positive entre éducation et revenu. L'investissement éducatif est plus rentable chez les femmes et dans les zones à faibles et moyens revenus. Le rendement éducatif moyen est aux alentours de 10%. La relation log-revenu-éducation n'est pas linéaire. La courbe du rendement de l'investissement éducatif prend généralement la forme de U à mesure qu'on avance d'un cycle éducatif à un autre. Le taux de rendement atteint son minimum au niveau des études secondaires et son maximum au niveau des études supérieures. Dans certains pays le rendement de l'investissement éducatif est plus élevé dans les zones urbaines et le secteur privé comparés à la zone rurale et au secteur public.

Cependant, la simplicité du modèle a un coût. Premièrement le modèle est développé sous des hypothèses dont la validité de certaines d'entre elles est en question. La méthode ne prend pas en compte les coûts directs de l'investissement éducatif ce qui peut biaiser les estimations. Le modèle défend l'idée que le revenu dépend uniquement du capital humain accumulé et néglige l'effet de l'origine sociale et de la discrimination. Ainsi le genre, l'appartenance ethnique et la couleur de la peau sont des variables manquantes. Les caractéristiques non observables posent également le problème d'endogénéité qui peut conduire à une surestimation du rendement éducatif. Toutes ces difficultés relevées nous amènent à se poser des questions à propos de la précision des estimations générées par l'application du modèle de Mincer tel qu'il est. Toutefois, ces défis empiriques ont généré des réflexions plus poussées qui ont abouti à des solutions et des corrections apportées au modèle classique. Les estimations sont probablement plus précises mais le train de l'amélioration doit continuer afin d'affronter des difficultés plus théoriques notamment la question des coûts directs.

## **Références**

- [1] G. S. Becker, «Investment in Human Capital: A Theoretical Analysis,» *Journal of Political Economy*, vol. 70, n° 5, pp. 9-49, 1962.
- [2] J. A. MINCER, SCHOOLING, EXPERIENCE, AND EARNINGS ; Individual Acquisition of Earning Power, NBER, 1974, pp. 5 - 23.
- [3] G. S. Becker et B. R. Chiswick, «Education and the Distribution of Earnings,» *The American Economic Review*, vol. 56, n° 1/2, Mar 1966.
- [4] J. A. Mincer, SCHOOLING, EXPERIENCE AND EARNINGS ; Schooling and Earnings, National Bureau of Economic Research / NBER, 1974, pp. 41 - 63.
- [5] G. S. Becker, Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, National Bureau of Economic Research, 1975, pp. 45 - 144.
- [6] G. Psacharopoulos et E. Velez, «EDUCATION AND THE LABOR MARKET IN URUGUAY ; the World Bank,» n° 28, 1992.
- [7] D. Lam et R. F. Schoeni, «Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil,» *Journal of Political Economy*, vol. 101, n° 4, pp. 710-740, 1993.
- [8] G. Psacharopoulos et F. Steier, «Education and the Labor Market in Venezuela 1975-1984,» *THE WORLD BANK, EDUCATION AND TRAINING SERIES, Discussion Paper*, 1987.
- [9] G. Psacharopoulos et E. Velez, «Schooling, Ability, and Earnings in Colombia, 1988,» *Economic Development and Cultural Change*, vol. 40, n° 3, pp. 629-643, 1992.

- [10] A. Aali-Bujari, F. Venegas-Martínez et A. García-Santillán, «Schooling levels and wage gains in Mexico,» *Economics and Sociology*, vol. 12, n° 4, pp. 74-83, 2019.
- [11] M. Aslam, F. Bari et G. Kingdon, «Returns to Schooling, Ability and Cognitive Skills in Pakistan,» *Research Consortium on Educational Outcomes and Poverty (RECOUP Working Paper)*, n° 20, 2008.
- [12] M. Asadullah et S. Xiao, «Labor Market Returns to Education and English Language Skills in the People's Republic of China: An Update,» *Asian Development Review*, vol. 36, n° 11, p. 80–111, 2019.
- [13] H. A. Patrinos, G. Psacharopoulos et A. Tansel, «returns to investment in education : the case of Turkey,» *Policy reserch Working paper 8789 ; Wold Bank Group ; Education Global Practice*, 2019 ;.
- [14] H. A. Patrinos et G. Psacharopoulos, «Returns to Investment in Education A Decennial Review of the global literature , Policy Research Working Paper,» *WORLD BANK GROUP*, n° 8402, pp. 1-24, April 2018.
- [15] C. E. Montenegro et H. A. Patrinos, «Comparable Estimates of Returns to Schooling Around the World (Policy Research Working Paper),» *World Bank*, n° 7020, 2014.
- [16] G. PSACHAROPOULOS, «Retums to Education: A Further International Update and hnplications,» *World Bank Reprint Series Reprinted with permission from The Journal of Human Resources*, vol. XX, n° 4, pp. 584-604, 1985.
- [17] O. ELKHALFI, R. CHAABITA et C. GUEMIMI, «Le rendement de l'enseignement supérieur par genre : Cas du Maroc,» *African Scientific Journal*, vol. 4, n° 3, pp. 648-662., 2021.
- [18] A. Tokila et H. Tervo, «Regional differences in returns to education for entrepreneurs versus wage earners,» *The Annals of Regional Science*, vol. 47, n° 3, pp. 689-710, 2011.
- [19] L. M. Brotherhood, P. C. Ferreira et C. Santos, «Education Quality and Returns to Schooling: Evidence from Migrants in Brazil,» *Escola Brasileira de Economia e Finanças (EPGE/FGV)*, n° 782, 2017.
- [20] A. El-Araby Aly et J. F. Ragan Jr., «Arab immigrants in the United States: how and why do returns to education vary by country of origin?,» *Journal of Population Economics*, vol. 23, n° 2, pp. 519-538, 2010.
- [21] N. Zaoujal et B. Outtaj, «Le rendement de l'éducation sur le marché du travail marocain : une analyse microéconométrique,» *Repères et Perspectives Economiques*, vol. 4, n° 2, pp. 265-282, 2020.
- [22] R. CHAABITA, O. ELKHALFI et C. GUEMIM, «L'effet établissement sur le rendement de l'enseignement supérieur au Maroc,» *Revue Française d'Economie et de Gestion*, vol. 2, n° 4, pp. 288-309, 2021.
- [23] H. ATI et K. SABRI, «Remise en question de la linéarité du rendement de l'éducation sur le salaire: Cas de la région de Casablanca,» *International Journal of Accounting, Finance, Auditing, Management and Economics*, vol. 3, pp. 322-336, 2022.
- [24] J. R. Behrman et N. Birdsall, «the quality of Sdooling quantity Alone is Misleading,» *World Bank Reprint Series: Number 376 ; Reprinted with permission from The American Economic Review*, vol. 73, n° 5, pp. 928-46, 1983.
- [25] J. Strauss et D. Thomas, «Wages, Schooling and Background: Investments in Men and Women in Urban Brazil,» *Economic Growth Center (EGC), Yale University , Center Discussion Paper*, n° 649, 1991.

- [26] K. TANGTIPONGKUL, «Rates of Return to Schooling in Thailand,» *Asian Development Review*, vol. 32, n° 2, pp. 38-64, 2015.
- [27] Z. Griliches et W. M. Mason, «Education, Income, and Ability,» *Journal of Political Economy*, vol. 80, n° 3, pp. 74-103, 1972.
- [28] J. R. Behrman et A. B. Deolalikar, «SCHOOL REPETITION, DROPOUTS, AND THE RATES OF RETURN TO SCHOOLING : THE CASE OF INDONESIA,» *OXFORD BULLETIN OF ECONOMICS AND STATISTICS*, vol. 53, n° 4, 1991.
- [29] J. J. Heckman, «Sample Selection Bias as a Specification Error,» *Econometrica*, vol. 47, n° 11, pp. 153-161, 1979.
- [30] C. U. CHISWICK, «ON ESTIMATING EARNINGS FUNCTIONS FOR LDCs,» *World Bank Reprint Series: Number Forty-four ; Reprinted from the Journal of Development Economics*, n° 3, pp. 67-78, 1976.