

Université de Montréal

**Le rendement de l'éducation au Canada en 1993:
une estimation par la méthode des variables instrumentales**

Présenté
par
Jean Philippe Bouchard

Sous la direction
de
Thomas Lemieux (Ph.D.)

Rapport de recherche présenté à la faculté des Études Supérieures en vue de l'obtention du
grade de Maître ès Sciences (M.Sc.) en Sciences Économiques

Département de Sciences Économiques
Faculté des Arts et des Sciences
Décembre 1999

Nous tenions a remercier M.Thomas Lemieux pour sa patience et sa disponibilité, ainsi que pour sa curiosité contagieuse.

Nos remerciements s'adressent également à M.Claude Montmarquette pour sa grande science et ses envolés.

Jean Philippe Bouchard, 1999

RÉSUMÉ

La corrélation entre l'éducation et le revenu a toujours existée. On observe ainsi souvent que les individus avec le plus d'années de scolarité sont ceux qui gagnent le salaire le plus haut.

Par ce travail, nous tentons de prouver que ce lien causal existe bien et nous estimons le rendement de l'éducation. Nous soulignons, par une synthèse des récentes études sur le sujet, que les estimations de ce rendement diffèrent selon la méthode d'estimation retenue : la méthode des moindres carrés ordinaires donne une valeur plus faible que celle obtenue par la méthode des variables instrumentales. Nous avons retenu la méthode des variables instrumentales pour corriger le biais d'habileté.

Notre estimation par les variables instrumentales se base sur les différences dans les systèmes scolaires canadiens. Notre modèle est appliqué à l'Enquête sur la Dynamique du Travail et du Revenu, une base de données canadiennes de 1993.

Nous obtenons un rendement de l'éducation de 8,5% par la méthode des variables instrumentales et de 6% par les MCO.

Le faiblesse du rendement de l'éducation peut s'expliquer par notre choix d'instrument. En effet, il ne vise pas des individus qui auraient un rendement marginal de l'éducation élevé.

“Il ne semble exister aucune limite à l’aptitude intellectuelle de l’homme à accroître ses connaissances scientifiques et à appliquer ses connaissances au progrès de sa technologie.”

-Arnold Toynbee

“Les penseurs, les érudits, les mathématiciens, les praticiens des sciences théoriques ou naturelles sont des êtres possédés. Sous la coupe d’une déraison dominante.”

-George Steiner

TABLE DES MATIÈRES

<u>Introduction</u>	p.1
<u>1.Revue de littérature</u>	p.2
1.1.Historique des recherches	p.2
1.1.1.Problématique	p.2
1.1.2.Les résultats théoriques	p.3
1.2.Les problèmes économétriques rencontrés lors des estimations	p.5
1.2.1.Le biais d'habileté	p.5
1.2.2.Les erreurs de mesures	p.7
1.3.Solutions économétriques	p.10
1.3.1.La méthode de l'estimateur à effet fixe	p.10
1.3.2.La méthode des variables instrumentales	p.11
<u>2.Modélisation</u>	p.17
2.1.Le modèle théorique	p.17
2.2.La méthode économétrique	p.18
2.3.L'instrument	p.23
<u>3.Les données</u>	p.24
3.1.L'Enquête sur le Dynamique du Travail et du Revenu	p.24
3.2.Analyse statistique	p.25
3.3.Construction des variables	p.27
<u>4.Les estimations</u>	p.32
4.1.Les résultats	p.32
4.1.1.Moindre Carré Ordinaire	p.32
4.1.2.Variable instrumentale	p.33
4.2.L'analyse des résultats	p.34
<u>Conclusion</u>	p.38
Annexe I	p.39
Annexe II	p.41
Bibliographie	p.42

SOMMAIRE DES TABLEAUX ET FIGURE

Tableau I : Synthèse des études marquantes utilisant la méthode des Variables Instrumentales

Tableau II : Moyenne d'année de scolarité par régions

Tableau III : Effets de l'instruction parentale sur le choix d'instruction de l'individu par régions

Tableau IV : Détail des variables utilisées dans la recherche

Tableau V : Fréquence croisée entre la région d'instruction et la région de résidence

Tableau VI : Moyenne d'année de scolarité par région et par catégorie d'instruction

Tableau VII : Résultats de l'estimation par les MCO

Tableau VIII : Résultats de l'estimation par les VI

Figure I : Les différences dans le nombre d'années de scolarités par régions par cohortes

INTRODUCTION

La corrélation entre l'éducation et le revenu a toujours existée. On observe ainsi souvent que les individus avec le plus d'années de scolarité sont ceux qui gagnent aussi le salaire le plus haut. Cependant, de la corrélation positive au lien causal, certains sont assez réticents à faire le pas. Ce n'est pas notre cas, puisque par ce travail, nous entendons prouver qu'il existe bien un lien causal et donner une estimation de sa valeur. Pour cela, nous nous intéresserons à la notion de rendement pur de l'éducation, i.e. à l'augmentation en pourcentage du salaire horaire associé à une année supplémentaire d'éducation.

Dans un premier temps nous ferons un rapide historique des recherches portant sur ce sujet et nous présenterons les problèmes généralement soulevés lors de l'estimation du rendement de l'éducation. Nous exposerons également les stratégies les plus fréquemment utilisées pour résoudre ces problèmes et les résultats qui s'y réfèrent. Dans un second temps, nous proposerons une nouvelle application de la stratégie des variables instrumentales. Nous pourrons alors élaborer notre modèle théorique et extrapoler de ses implications. La dernière section sera réservée à l'analyse des résultats obtenus.

1. Revue de littérature

1.1. Historique des recherches

1.1.1. Problématique

Nous savons depuis longtemps que les caractéristiques familiales jouent un rôle important dans le développement cognitif de l'enfant. Elles peuvent ainsi influencer son intérêt pour le monde extérieur et sa curiosité, la qualité de l'éducation qu'il reçoit, et la quantité d'informations qu'il acquière en une année. L'éducation d'un individu sera ainsi influencée directement par l'éducation qu'ont reçue ses parents (exposition précoce aux connaissances) et par ce qu'elle implique (revenus supérieurs, emplois plus intéressants, meilleurs réseaux professionnels, etc...). Plusieurs auteurs ont évalué l'effet de l'éducation des parents sur le choix de scolarité des enfants, qui atteint 0.5 pour Altonji et Dunn (1996) et 0.6 pour Ashenfelter et Rouse (1998). Sans pour autant nous avancer sur la valeur numérique de cet effet, nous reconnaissons son existence. Dans plusieurs recherches, on utilise l'effet de l'éducation des parents sur le choix d'éducation des enfants pour instrumenter la variable d'éducation¹. On peut cependant utiliser d'autres instruments, par exemple l'existence de réformes institutionnelles ou la proximité des institutions scolaires du lieu de résidence (Card).

En raison de la non-homogénéité des systèmes scolaires au Canada, de par leur détermination au niveau provincial, nous avons choisi d'utiliser l'effet "scolarité des parents" dans l'estimation du rendement de l'éducation. De plus, nous supposons, dans l'élaboration de notre variable instrumentale, que les individus au Canada sont homogènes. L'utilisation de la méthode des variables instrumentales répond au besoin de corriger le biais d'habileté. Avant d'aller plus loin

¹ Voir Griliches (1979), qui fait une très bonne recension de la littérature utilisant l'effet de l'éducation des parents sur le choix de scolarité des enfants

sur ce sujet, nous allons présenter un éventail des recherches les plus marquantes dans ce domaine, ainsi que les théories nécessaires à une bonne compréhension des nuances des recherches présentées.

1.1.2. Les résultats théoriques

Ce sont les penseurs de la théorie du Capital Humain qui en premier ont avancé l'idée du lien causal entre l'éducation et le revenu (Becker, 1964). En tentant d'expliquer l'origine de la croissance des États-Unis après la Deuxième Guerre Mondiale, ils ont soumis l'hypothèse que l'augmentation de la scolarité de la population pouvait expliquer en grande partie la croissance économique observée au États-Unis. Cette hypothèse ne fut pas acceptée par toute la communauté scientifique. Les critiques faisaient observer que le lien causal entre éducation et revenu ne pouvait exister que dans le cas où écarts salariaux traduisaient écarts de productivité et non un écart d'habileté, potentiellement corrélé avec l'éducation. Cependant, cette critique peut être comprise autrement. Les individus plus habiles ont deux avantages : ils sont plus productifs et ils ont plus de facilité à acquérir des connaissances. Ainsi, ils peuvent rester à l'école plus longtemps, à moindre coût, ce qui résulte en un nombre d'années d'éducation plus élevé et donc un salaire supérieur. C'est de cette explication qu'est née la Théorie du Signal (Spence). Notons pour l'anecdote que Becker (1964), s'interrogeant sur ces critiques, déclara que si les chercheurs accordaient autant d'importance à l'habileté, c'est qu'ils se croyaient plus habiles que la moyenne ayant eux-mêmes une éducation plus grande que la moyenne, une sorte d'excès d'humilité en quelque sorte...

Reste que la réalité se situe sûrement entre ces deux extrêmes : l'habileté s'acquiert avec l'éducation et/ou l'éducation s'acquiert avec l'habileté. C'est de cette interrogation que naîtra toute une série de problèmes reliés à l'estimation du rendement de l'éducation.

1.2. Les problèmes économétriques rencontrés lors des estimations.

1.2.1 Le biais d'habileté

L'existence du lien causal entre l'éducation et le revenu est de plus en plus acceptée, et les preuves tendant à le démontrer se font de plus en plus nombreuses². Mais cet avis n'est pas accepté par tous, certains s'appuyant sur l'existence d'un "biais d'habileté" pour la réfuter. Or, ce biais, comme nous le présenterons tout au long de ce travail, est une notion qui ne peut être séparée de l'estimation du rendement de l'éducation.

Pour expliquer ce qu'est le biais d'habileté, nous prenons à Card (1995a) son modèle de choix de l'éducation et de détermination de revenu :

$$1- \quad \text{Log } y_i = X_i \beta + S_i \rho + \mu_i$$

$$2- \quad S_i = X_i \gamma + \nu_i$$

où S est une mesure du nombre d'années d'éducation, y est une mesure du revenu, X un vecteur de caractéristiques personnelles observables et μ et ν sont des résidus. Le rendement de l'éducation est représenté par le coefficient ρ .

L'estimation de ρ par les moindres carrés ordinaires ne pose pas de problèmes dans la mesure où μ et ν ne sont pas corrélés. Cependant, dans notre cas, l'estimation par MCO donne des estimateurs biaisés et non consistants de l'effet causal des variables explicatives (éducation) sur le résultat (Bound, Jaeger et Baker, 1995).

² Voir Ashenfelter et Rouse, 1999

Ce biais d'habileté résulte de la présence dans le terme d'erreurs de caractéristiques non observables de l'individu telles que son habileté ou sa motivation, qui peuvent influencer aussi bien le nombre d'années d'éducation (on dira que la variable S est endogène) que le revenu. Étant incapable de les inclure dans les équations, on les retrouve dans les termes d'erreurs μ et ν , qui sont donc corrélés.

Le modèle utilisé par Griliches (1977) illustre bien ce biais.

$$1' \quad \log Y = \alpha + \beta S + \gamma A + \mu$$

où A est l'habileté non observable et μ un terme d'erreurs. On suppose que seule l'habileté biaise les estimateurs, ce qui revient à dire que μ n'est pas corrélé avec les variables explicatives.

Lorsqu'on omet A dans la régression par MCO de l'équation 1' on trouve:

$$E [b] = \beta + \gamma c$$

où b est l'estimation du rendement de l'éducation, γ est l'effet de l'habileté sur le revenu et c est l'effet de l'éducation sur l'habileté.

Notre estimateur est donc biaisé vers le haut.

Cette conclusion générale nécessite toutefois le postulat de trois hypothèses :

- i) l'habileté a un effet positif direct sur le revenu ($\gamma > 0$) et plus grand que celui sur l'éducation *correctement* mesurée
- ii) l'effet de l'éducation sur l'habileté est positif ($c > 0$)
- iii) c'est la seule variable omise et toutes les autres hypothèses relatives au MCO tiennent

1.2.2. Les erreurs de mesures

Selon Griliches (1977), les recherches sur le rendement de l'éducation portent une attention excessive au biais positif occasionné par le problème du biais d'habileté ou d'omission de variables autres. Ce comportement mène à des biais négatifs qui peuvent être très importants si la mesure de l'éducation n'est pas parfaite ou encore si elle est sujette à des erreurs de mesures aléatoires. Par exemple, lorsqu'on effectue une enquête, on s'expose toujours au risque que la réponse fournie soit fautive ou imprécise, et ce quelle que soit la question posée. Ainsi, Ashenfelter et Rouse (1998) estiment à 90 % la fiabilité des données portant sur le nombre d'années d'éducation. La marge d'erreur semble donc faible. Cependant, elle ne l'est pas en réalité, puisqu'elle est multipliée lors de l'ajout de variables contrôlant l'éducation. Ainsi, plus nous ajoutons de variables reliées à l'éducation, plus les erreurs de mesures prennent une place importante dans le biais. *“ The amount of information contained in any one specific data set is finite and, therefore, as we keep asking finer and finer questions, our answers become more and more uncertain ”*³

Ainsi, pour Griliches (1977), le biais négatif associé aux erreurs de mesures vient annuler l'effet du biais d'habileté et des variables omises. Il peut même arriver selon l'auteur que le biais négatif soit supérieur au biais d'habileté. En reprenant les exemples pré-cités, on trouverait pour l'étude de Card une valeur estimée du rendement de l'éducation b par MCO non biaisée, et pour l'étude de Griliches, une valeur de b biaisée vers le bas.

³ GRILICHES, Zvi (1977)

Cependant ce résultat, qui faciliterait de beaucoup l'estimation, ne tient que dans le cas où les erreurs de mesures ne sont pas corrélées avec le *vrai* niveau d'éducation.

Or, pour Card (1998), les erreurs de mesures sont corrélées négativement avec le vrai niveau d'éducation. En effet, lors du questionnement des individus, ceux avec peu d'éducation n'ont pas intérêt à donner de réponses inférieures à leur vrai niveau de scolarité, l'inverse étant vrai pour les individus avec un haut niveau d'éducation qui ne peuvent fournir de réponses plus grandes. De plus, Card estime comme Ashenfelter et Rouse (1998) que la fiabilité des données est légèrement supérieure à 90 %. Il conclue que le biais des erreurs de mesures est ou légèrement inférieur à celui de l'habileté ou identique.

En conclusion, nous pouvons réunir ces résultats dans un tableau simple :

GRILICHES	$\phi \leq \rho$
CARD	$\phi \geq \rho$

Où ϕ : biais d'habileté

ρ : biais des erreurs de mesures

Le problème du biais d'habileté nous empêche de répondre clairement à la question : "existe-t-il un lien causal entre l'éducation et les revenus, autrement que par les effets des variables omises?"

Un moyen efficace mais peu réalisable pour le faire serait de prendre deux groupes d'individus identiques en tous points en moyenne, et de leur donner un niveau d'éducation différent. Ainsi, les individus du groupe "témoin" ne recevraient pas d'éducation, et serviraient ainsi de point de

comparaison avec ceux du groupe du traitement qui eux recevraient un niveau de scolarité plus élevé.

Les deux groupes seraient créés de manière aléatoire, ce qui éliminerait le problème d'habileté, les individus étudiants n'étant pas forcément les plus habiles. A l'issue de l'expérience la seule différence entre les étudiants en moyenne serait le niveau d'éducation. On pourrait alors relever les différences salariales et en tirer le rendement *pur* de l'éducation, non biaisé.

Cette méthode n'est pas réalisable sur un plan éthique, car on ne peut choisir subjectivement de priver des individus de s'éduquer. Cependant, on peut noter pour l'anecdote qu'Ashenfelter et Rouse (1999) estiment que ce type d'expérience ne soulève pas de problèmes moraux si on la réalise dans un pays du tiers monde, argumentant que les individus de ces pays n'obtiendraient de toutes façons pas d'éducation...

Plusieurs solutions économétriques ont été avancées pour régler les problèmes que nous venons de mentionner. Card (1995) fait une distinction entre deux sortes de méthodes : la méthode des variables instrumentales⁴ et la méthode de l'estimateur à effet fixe.

⁴ Cette méthode est détaillée de manière exhaustive dans un article de Angrist, Imbens et Rubin (1996).

1.3. Solutions économétriques

1.3.1. La méthode de l'estimateur à effet fixe.

Cette stratégie prend pour hypothèse que les résidus des équations 1 et 2 ont une structure de covariance restrictive. Une fois cette hypothèse établie, un chercheur doit utiliser :

- i) des observations répétées dans le temps pour un individu (panel) ou
- ii) des observations de différents individus de la même famille pour enlever la corrélation des résidus

C'est cette méthode qui est utilisée dans les recherches sur les jumeaux, les plus connues étant celles d'Ashenfelter et Rouse (1998). Elle se fonde sur l'hypothèse que les membres d'une même famille ont essentiellement la même habileté, ce qui est plausible si on admet que l'habileté est innée et génétiquement transmissible. Dans ce cas, deux frères ayant une scolarité identique devraient avoir le même revenu. Les études peuvent ainsi porter sur des couples parent-enfant, enfant-enfant ou sur des jumeaux (monozygote ou dizygote).

Les études les plus marquantes sont celles sur les jumeaux. Les jumeaux *identiques* ont une composition génétique identique. De plus, ils grandissent et évoluent dans le même milieu, et pour tant soit peu que les parents ne les traitent pas de façon complètement différente, ils ont potentiellement une habileté identique. En contrôlant ainsi pour l'habileté, on ne peut plus justifier par des différences d'habileté une corrélation entre le fait que le jumeau le plus éduqué soit aussi celui qui gagne le plus d'argent.

Dans leur étude, Ashenfelter et Rouse (1998) interrogent des jumeaux *identiques* sur leur propre revenu et éducation, puis demandent à chacun des jumeaux de donner la scolarité de son frère. Ils utilisent cette dernière information pour contrôler les erreurs de mesures dans les réponses.

Comme on peut le voir dans le tableau I, l'estimateur des MCO pour le rendement de l'éducation est alors d'environ 8%, ce qui signifie qu'une année supplémentaire d'éducation pour un des jumeaux implique une différence de 8% dans les revenus des jumeaux. De plus, si on ne tient pas compte de leur situation de jumeaux, le rendement de l'éducation passe à 10,6%, soit un biais positif de 25 %. Ainsi, il semblerait que le biais d'habileté est d'environ 25%. Enfin, en instrumentant pour corriger les erreurs de mesures, ils obtiennent un rendement de l'éducation de 10%. On peut donc penser que le biais des erreurs de mesures vient effectivement corriger le MCO vers le bas.

1.3.2. La méthode des variables instrumentales.

Pour mieux apprécier l'étendue des instruments utilisés pour estimer le rendement de l'éducation, voici un bref aperçu des articles les plus marquants où les auteurs utilisent la méthode des variables instrumentales pour corriger le biais des erreurs de mesure. Notons qu'Ashenfelter et Rouse (1999) et Card (1995a) en font une très bonne synthèse.

Tous les résultats mentionnés seront repris à la fin de cette section dans un tableau de synthèse.

Angrist et Krueger (1991) (AK-91) se servent du *trimestre de naissance des individus nés dans les années 30, 40 et 50* comme variable instrumentale. Ils arrivent au résultat que le nombre

d'année d'éducation dépend du trimestre de naissance : les individus nés dans le premier trimestre ont en moyenne une année scolaire de moins que les autres. En effet, les individus nés dans le premier trimestre vont être dans la même classe que ceux nés dans les trois autres trimestres. Cependant, la loi sur la scolarité obligatoire permet de quitter l'école à 16 ans. Comme ceux qui sont nés dans le premier trimestre vont remplir cette condition avant ceux nés dans les trois autres trimestres, et ils vont quitter l'école en ayant à leur actif une année de scolarité de moins.

En prenant pour hypothèse que le trimestre de naissance n'a pas d'influence direct ni sur les revenus, ni sur l'habileté, on peut se servir du trimestre comme variable instrumentale. Les résultats obtenus sont intéressants : les individus nés dans le premier trimestre (first born) ont en moyenne un revenu annuel inférieur aux autres de 1%. Ce qui donne un rendement de l'éducation d'environ 8%. Ashenfelter et Rouse (1998) mentionne pour la même étude un rendement de 10%.

Bound, Jaeger et Baker (1995), dans un article critiquant l'usage abusif des variables instrumentales démontrent que lorsque l'instrument n'est que faiblement corrélé avec les erreurs dans les équations structurelles 1 et 2⁵, l'estimateur des VI est biaisé dans la même direction que l'estimateur des MCO. La critique à l'égard de AK-91 porte essentiellement sur la faiblesse de la corrélation entre le trimestre de naissance et l'éducation, ce qui pourrait laisser croire que leur rendement de l'éducation estimé est biaisé vers le haut. Cette critique de Bound, Jaeger et Baker est formulée envers un autre article où Angrist et Krueger utilisent la Guerre du Vietnam comme instrument.

⁵ Voir page 5

Dans cet article, Angrist et Krueger (1992) utilisent *les numéros tirés lors de la loterie du Vietnam* (Vietnam Draft Lottery) pour instrumenter les années d'éducation et le statut de vétéran. Cette loterie se déroulait de la manière suivante. Chaque date de naissance correspondait à un numéro allant de 1 à 365. Lors d'une émission télévisée, on tirait tour à tour un des 365 numéros, et les individus dont la date de naissance correspondait aux premiers numéros tirés partaient pour la Guerre. Cependant, certains individus étaient exemptés, par exemple ceux qui étaient inscrits dans les Universités comme étudiants. Cette offre d'exemption eut d'ailleurs l'effet d'augmenter le taux d'inscription dans les institutions américaines pour les individus dont la date de naissance correspondait aux premiers numéros tirés.

Tous ces individus exemptés pour cause scolaire ayant été choisis de façon aléatoire (la loterie) et ayant une éducation supérieure à la moyenne, il était possible de calculer le rendement de l'éducation en observant les différentiels de salaire et d'éducation. Les auteurs contrôlaient également pour le statut de vétéran, catégorie d'individus sensiblement différente des autres. Dans cette étude, les résultats obtenus par la méthode des variables instrumentales est de 10% supérieure au rendement estimé par les MCO.

Card (1995b) utilise *la proximité entre le lieu de résidence et l'université* comme variable instrumentale. Il démontre que le fait de résider à proximité d'une université influence le choix du nombre d'année d'éducation sans influencer directement les revenus. Le rendement de l'éducation obtenu par la méthode des VI est de 80% supérieur à celui des MCO. Mais de l'aveu de Card, cet effet est concentré principalement chez les individus de familles défavorisées, familles pour qui les coûts de transport ou de logement d'un enfant allant à l'université seraient

trop élevés. Card tient donc compte de ce biais en utilisant comme instrument l'interaction entre la proximité du lieu de résidence et les caractéristiques familiales. Il obtient quand même des résultats avec les VI de 30% supérieur à ceux des MCO.

Une autre approche originale nous vient de Card et Lemieux (1998), qui utilisent *l'éligibilité potentielle au Programme de Réhabilitation des Vétérans (VRA)* comme variable instrumentale. Ce programme offrait au vétérans de la Deuxième Guerre Mondiale qui acceptaient de retourner aux études certains avantages, notamment financiers. En étudiant les statistiques recueillies concernant ce programme, on s'aperçoit que son impact a été très différent selon les provinces dans lesquelles il a été mis en œuvre. Le programme a ainsi eu un effet nettement plus important en Ontario qu'au Québec, entre autre parce que le système scolaire québécois ne facilitait pas la chose et parce que moins de Québécois avaient le statut de vétérans. Ainsi, pour la cohorte visée en Ontario, le programme a résulté en 0,3 années d'éducation de plus que s'il n'avait pas existé. Les auteurs trouvent par la méthode des VI un rendement de l'éducation d'environ 15%, et par la méthode des MCO un rendement de 7%. Le biais des erreurs de mesures et une certaine hétérogénéité du rendement de l'éducation peuvent expliquer cet écart. Nous reviendrons sur cet article dans la section suivante pour expliquer le contexte scolaire canadien avant 1950.

Le tableau suivant reprend les résultats empiriques des études que nous venons de présenter.

Tableau I : Synthèse des études marquantes utilisant la méthode des Variables Instrumentales

ETUDES	STRATÉGIE	MÉTHODE	OLS VI		
ASHENFELTER ET ROUSE 1998	EFFET FIXE +VARIABLE INSTRUMENTALE	JUMEAUX IDENTIQUES	7.80%	9.90%	Écart-type
			0.017	0.021	
ANGRIST ET KRUEGER 1991a	VARIABLE INSTRUMENTALE	TRIMESTRE DE NAISSANCE	6.30%	8.10%	Ecart-type
			0.000	0.011	
ANGRIST ET KRUEGER 1992	VARIABLE INSTRUMENTALE	LOTERIE DU VIETNAM	5.90%	6.60%	Ecart-type
			0.001	0.015	
CARD 1995b	VARIABLE INSTRUMENTALE	PROXIMITÉ D'UNE UNIVERSITÉ AVEC LE LIEU DE RESIDENCE	7.30%	13.20%	n-corrigé
			0.004	0.049	Ecart-type
				9.70%	corrige
			0.005	Ecart-type	
LEMIEUX ET CARD 1998	VARIABLE INSTRUMENTALE	VETERAN REHABILITATION ACT	7%	14%	Ecart-type
			0.002	0.091	

Ces cinq articles arrivent sensiblement à la même conclusion : l'estimation du rendement de l'éducation par les VI est supérieure à celle obtenue par les MCO. Ce résultat peut paraître surprenant parce que l'estimation par les VI est supposée corriger le biais positif de l'estimation par MCO. Quatre explications à ce phénomène ont été trouvées :

1. Bound et Jaeger (1995) avancent que l'estimateur des VI est encore plus biaisé vers le haut que celui des MCO en raison de différences non-observées entre les caractéristiques des groupes de traitements et de contrôle, reprenant ainsi leur idée que les instruments utilisés habituellement ne sont que faiblement corrélés avec la variable d'éducation. Cette faiblesse biaiserait l'estimateur des VI dans le même sens que celui des MCO.

2. Griliches (1977) affirme que le biais d'habileté dans l'estimateur des MCO est relativement petit en comparaison d'autres biais, notamment celui des erreurs de mesures. Cela signifie que l'estimateur des MCO n'est pas biaisé, ou peut l'être, mais de façon non significative.
3. Ashenfelter et Harmon (1998) ont trouvé un biais pour le moins original : le biais de publication. Les chercheurs voulant trouver un nouvel estimateur des VI prendront celui qui donne le résultat le plus important, dans le sens d'élevé, pour le rendement de l'éducation.
4. Card (1998) revient sur un des points de son article sur la proximité scolaire. Les instruments choisis comme l'éducation obligatoire ou la proximité des universités affectent de manière plus significative les personnes qui n'auraient pas été à l'école autrement, des personnes qui ont un rendement marginal de l'éducation plus élevé que la moyenne. C'est donc pourquoi l'estimation par les VI donnera comme résultat un rendement de l'éducation supérieur à celui des MCO. Bref, il y a un problème d'hétérogénéité du rendement de l'éducation.

Maintenant que les bases théoriques, historiques et pratiques sont posées, nous pouvons passer à la partie pratique de ce travail : l'estimation du rendement de l'éducation au Canada.

2. Modélisation

2.1. Le modèle théorique

Modèle avec deux régions (1,2)

$$1- \log y_{i1} = c_1 + \beta X_i + \lambda S_i + \pi F_i + \eta_i$$

$$2- \log y_{i2} = c_2 + \beta X_i + \lambda S_i + \pi F_i + \eta_i$$

$$3- S_{i1} = b_1 + \xi X_i + \phi_1 F_i + \varepsilon_i$$

$$4- S_{i2} = b_2 + \xi X_i + \phi_2 F_i + \varepsilon_i$$

où y_i est une mesure du salaire horaire moyen, X les caractéristiques individuelles, S le nombre d'années d'éducation, F l'éducation des parents en année et η un terme d'erreur. Nous prenons comme hypothèse que la scolarité des parents ne peut servir d'instrument parce qu'elle influe directement sur les revenus ($\pi \neq 0$).

L'estimation que nous ferons du rendement de l'éducation se base sur la différence entre ϕ_1 et ϕ_2 ($\phi_1 \neq \phi_2$). Le niveau moyen de scolarité étant différent selon les régions, l'effet que le diplôme aura sur la scolarité de l'enfant ne sera pas le même selon la région de résidence.

En combinant l'équation 1 avec l'équation 2, et l'équation 3 avec l'équation 4, nous obtenons respectivement les équations suivantes:

$$5- \log y_i = c_1 + \delta_i (c_2 - c_1) + \beta X_i + \lambda S_i + \pi F_i + \eta_i$$

$$6- S_i = b_1 + \delta_i (b_2 - b_1) + \xi X_i + \phi_1 F_i + \delta_i (\phi_2 - \phi_1) F_i + \varepsilon_i$$

Avec δ une variable dichotomique qui prend la valeur 1 quand l'individu réside dans la région 2 et 0 quand il réside dans la région 1.

2.2. La méthode économétrique

Dans la régression de ces équations simultanées, nous utilisons la méthode des variables instrumentales pour corriger le biais lié aux erreurs de mesures. L'instrument que nous utilisons est l'interaction entre la scolarité des parents et la région de résidence. Nous prenons comme hypothèse que ϕ_1 est différent de ϕ_2 ⁶.

Le choix de la méthode des variables instrumentales repose sur des faits historiques. Le cas du Canada présente une caractéristique qui s'avère pratique pour notre estimation : la co-existence de deux entités culturelles et/ou linguistiques distinctes.

Outre leur langue, les individus sont en moyenne identique. Plus précisément les individus né en Ontario, au Québec, dans les provinces de l'Ouest ou dans les Maritimes ne présentent en moyenne aucune différence d'habileté⁷.

Cependant, la différence linguistique traduit des systèmes scolaires différents. Ainsi, le système des collèges classiques faisait preuve de discrimination à l'admission⁸ dans les régions de l'est du pays alors que le système scolaire de l'ouest était plus démocratique.

⁶ Voir Section 2.3

⁷ Les quatre régions mentionnées seront celles utilisées dans les régressions ultérieures, mais à des fins de simplification, nous n'utiliserons que le Québec et l'Ontario pour décrire le modèle.

⁸ Pour être admis, les individus devaient être de classe moyenne ou se préparer à devenir prêtres.

L'étude du nombre d'années de scolarité en moyenne dans les différentes provinces canadiennes permet de mettre en lumière des disparités potentielles. Pour faciliter cette analyse, nous avons regroupé les provinces canadiennes en 4 catégories :

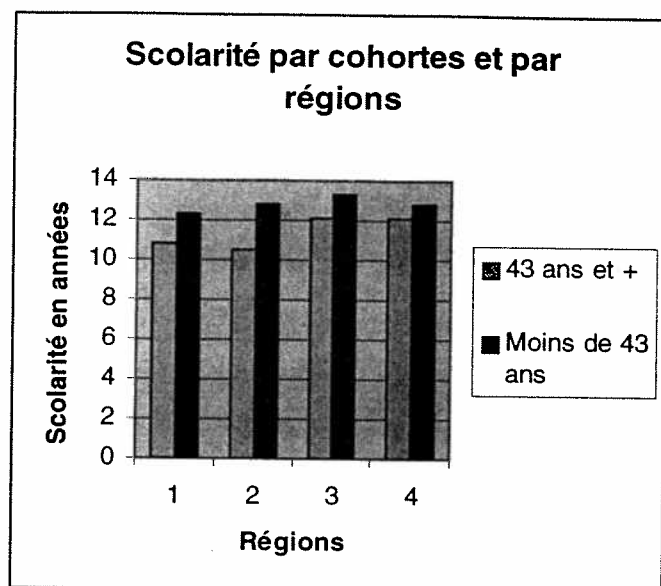
1. ATLANTIQUE : Nouvelle-Écosse, Nouveau-Brunswick, Terre-Neuve et Île-du-Prince-Édouard.
2. QUÉBEC
3. ONTARIO
4. OUEST : Manitoba, Saskatchewan, Alberta, Colombie-Britannique.

Les statistiques sont celles d'individus âgés entre 16 et 65 ans.

Tableau II : Moyenne d'année de scolarité par régions

	Année de scolarité en moyenne	
	43 ans et +	Moins de 43
Atlantique	10.7835138	12.2695288
é.-t.	3.3563412	2.7121712
Québec	10.499697	12.8443316
é.-t.	3.8972265	3.1258067
Ontario	12.1026444	13.3234621
é.-t.	3.2625366	2.6589153
Ouest	12.0638402	12.7864777
é.-t.	3.1783548	2.4179668

Figure I : Les différences dans le nombre d'années de scolarités par régions par cohortes



Le tableau 2 reprend le nombres d'années de scolarité dans chacun des groupes définis ci-dessus pour deux classes d'âge : les hommes et les femmes âgés de 43 ans et plus et ceux âgés de moins de 43 ans. Le choix de division pour les deux cohortes correspond à l'année où le système de collèges classiques a pris fin, en 1965. Les individus de 43 ans en 1993 sont ainsi ceux qui ont connus en dernier ce système. Cette séparation nous permet également d'obtenir deux cohortes de tailles similaires.

On peut remarquer deux tendances dans ce tableau.

En premier lieu, on remarque un écart visible entre l'est et l'ouest en terme d'années de scolarité pour le groupe de 43 ans et plus, le groupe de l'ouest dominant l'est. Cet écart est attribuable en partie à la difficulté d'intégrer les écoles dans l'est du pays et au système des collèges classiques, ce qui ne favorisaient pas la démocratisation du savoir.

Dans un second temps, qui correspond à la démocratisation de l'enseignement dans l'est du pays issu de l'abandon du système des collèges classiques, l'écart entre les deux régions tend à diminuer. La scolarisation s'homogénéise au sein de la population du Canada.

Maintenant que nous savons que les Canadiens sont homogènes à la naissance, nous pouvons penser que les différences salariales entre les individus des différentes provinces peuvent provenir des différences dans leurs niveaux respectifs de scolarité. En effet, on ne peut attribuer ces différences à des niveaux d'habileté. En quelque sorte, nous sommes dans le cas de l'Expérience décrite dans la section 1.2.2. où on évaluait le rendement de l'éducation en donnant une éducation différente à deux groupes identiques.

Par contre, des différences entre les régions telles que le climat nous empêchent d'utiliser la région de résidence et/ou d'instruction comme instrument.

En effet, selon la théorie des différences compensatoires, deux individus identiques, ayant le même emploi pour la même compagnie et ayant connus le même système d'éducation, les mêmes parents, etc...peuvent souffrir de différences salariales pour autant peu qu'ils travaillent dans deux villes différentes. Selon cette théorie, les écarts salariaux ne représentent pas uniquement des écarts de productivités (Rosen, 1986) mais reflètent également un arbitrage nécessaire pour *égaliser* les avantages (ou désavantages) monétaires ou non monétaires totaux parmi les branches d'activités ou parmi les travailleurs eux-mêmes. Cette compensation peut alors prendre place dans un cadre géographique entre deux régions ou deux villes. Des facteurs déterminants les différences compensatoires seront par exemple les différences dans le climat, le taux de criminalité, le taux de pollution ou la population.

Le choix de la variable "éducation des parents" comme instrument aurait pu être pertinent, d'une part parce qu'elle reflète bien les différences dans la qualité du système d'éducation et d'autre part parce que les différences d'habileté sont plus rares entre enfants et parents qu'entre voisins par exemple. Cependant, l'imparfaite mobilité générationnelle, qui fait que la position des parents sur le marché du travail peut biaiser la position des enfants (par exemple en aidant ces derniers à trouver un emploi dans une entreprise dirigée par des amis de la famille), rend ce choix non pertinent.

Ainsi, la région de résidence et/ou d'instruction et l'éducation des parents ne peuvent être choisies comme instrument car elles ont une influence *directe* sur le revenu des individus. Par contre, une variable d'interaction entre la région de résidence de l'individu et le niveau de scolarité du père influencera la scolarité de l'individu sans influencer directement son revenu. En effet, rien ne nous indique qu'une variable d'interaction peut influencer directement le revenu de l'individu, cette variable indiquant qu'il habite dans telle région et que ses parents ont étudié tant d'années. L'hypothèse principale de ce travail est donc que notre variable d'interaction n'a pas d'influence directe sur les revenus de l'individu.

2.3. L'instrument

La pertinence de l'instrument repose sur la différence entre ϕ_1 et ϕ_2 , ou de manière plus générale, les différences entre les effets de l'instruction parentale sur l'instruction individuelle par régions. Nous avons effectué les régressions permettant de quantifier ϕ_1 et ϕ_2 . En effectuant des MCO par région, nous avons obtenu les résultats suivants :

Tableau III : Effets de l'instruction parentale sur le choix d'instruction de l'individu par régions

Régions	Effet
Atlantique	1.079591
Québec	0.996677
Ontario	0.807886
Ouest	0.671471

Comme on s'y attendait, l'effet est plus fort en se dirigeant vers l'est du pays. On peut comprendre ce phénomène à la lumière des résultats présentés dans le Tableau II. Les niveaux de scolarité étant plus élevés dans l'ouest du pays, l'effet du niveau d'éducation des parents sera moindre. Le niveau d'instruction des parents aura moins d'effets sur celui de leurs enfants à mesure que le niveau moyen d'instruction s'élèvera. Notons que les résultats complets des régressions présentées dans le Tableau III sont fournies en Annexe I.

On remarquera dans ces résultats que la variable SEX prend une valeur négative dans les régions de l'est et de l'ouest. La variable étant une dichotomique prenant la valeur 1 quand l'individu est de sexe masculin, on peut en conclure qu'il y a corrélation négative entre masculinité et scolarité. Le fait étant d'actualité, nous trouvons important de le mentionner.

3. Les données

3.1. L'Enquête sur la Dynamique du Travail et du Revenu

Nous avons utilisé pour estimer le modèle précédent une base de données de Statistiques Canada qui possédait toutes les informations nécessaires à l'application de la méthode des variables instrumentales.

“L'Enquête sur la dynamique du travail et du revenu (l'EDTR) est une enquête longitudinale qui est effectuée auprès des ménages par Statistique Canada. Elle vise à retracer les changements qui touchent la situation économique des personnes et des familles, au fil des ans, ainsi que les facteurs qui ont une influence déterminante à cet égard. Les personnes choisies au départ pour l'enquête sont interviewées une ou deux fois par année, pendant six ans, en vue de recueillir des données sur leur expérience sur le marché du travail, leur revenu et leur situation familiale. Pour obtenir des renseignements complets sur les familles, ainsi que des données transversales, les personnes qui vivent avec les répondants originaux, à un moment ou à un autre au cours de la période de six ans, sont aussi interviewées au cours de cette période.”⁹

Pour l'estimation, nous utiliserons seulement les données de 1993. Les régressions ont été effectuées pour les données de 1994, mais ne présentent pas de différences significatives avec celles de 93, c'est pourquoi nous n'avons pas jugé bon de les présenter.

⁸ Texte tiré du Guide de l'Utilisateur de l'EDTR, Statistiques Canada.

3.2. Analyse statistique

La présentation des variables se fait en deux temps. Le tableau 3 reprend les variables de l'EDTR utilisées dans cette étude et le tableau 4 présente les variables qui ont été créées à partir de variables de l'EDTR. Les statistiques représentent les individus de tous âges, sexes, niveaux de scolarité et régions de résidence.

Tableau IV¹⁰ : Détail des variables utilisées dans la recherche

Variable	Moyenne	Ecart-type	Mini	Maxi	N
Age	37.3139829	13.1731487	16	65	22928
Sex	0.4916696	0.4999415	0	1	22928
Statmar	0.6384891	0.4804486	0	1	22926
Langmat	0.2659501	0.441848	0	1	21583
Minvis	0.0137519	0.1164617	0	1	22906
Edpere	2.0416991	1.2577873	1	5	20528
Edmere	2.1417943	1.1957991	1	5	20699
Immigra	0.0343001	0.1820029	0	1	22653
Revhora	13.8791124	7.3822904	2.09	61.31	14646
Exp	12.9227213	11.9156962	0	51	18632
Dicoedu	2.2056001	1.4631872	1	8	22928
Provedu	31.4704292	14.5550274	10	59	22928
Edu	12.3700285	3.0599112	1	20	22468
Catedu	7.0555751	2.5148335	2	12	22708
Habiatl	0.2267969	0.1487693	0	1	22928
Habique	0.2066905	0.4049404	0	1	22928
Habiont	0.2582868	0.4377021	0	1	22928
Habioue	0.3082258	0.4617704	0	1	22928
Eduatl	0.2321179	0.4221931	0	1	22928
Eduque	0.2158496	0.4114194	0	1	22928
Eduont	0.2621685	0.4398234	0	1	22928
Eduoue	0.2898639	0.4537089	0	1	22928

¹⁰ Les définitions des variables de Statistiques Canada sont regroupées dans l'annexe II.

Nos données représentent bien la population canadienne, si ce n'est que pour la statistique des minorités visibles (MINVIS) et celle de l'immigration (IMMIGRA). Elle nous indiquent que seulement 1% de la population est décrit par le gouvernement comme minorité visible et que 3% de la population se considère immigrante alors que les véritables statistiques de ces variables sont respectivement de 9.8% et de 17%¹¹. Ces variables ne représentant pas bien la population, nous ne les garderons pas afin de ne pas biaiser les résultats.

La variable SEX a été modifiée pour qu'elle reflète les avantages salariaux relatifs aux hommes. C'est-à-dire que la variable prend la valeur 0 quand l'individu est de sexe féminin ou indéterminé et 1 quand il est de sexe masculin. La statistique semble bien représentative de la population où environ des individus sont de sexe masculin.

L'indicateur de langue maternelle (LANGMAT) est une variable dichotomique qui prend la valeur 1 quand l'individu a comme langue maternelle le français. Elle représente bien la situation au Canada où les Québécois francophones et les francophones hors Québec comptent pour un quart de la population totale canadienne.

Les variables EDPERE et EDMERE donnent la scolarité des parents de l'individu. Cette variable peut prendre des valeurs de 1 à 5. La moyenne est de 2 aussi bien pour les hommes que pour les femmes, ce qui correspond au niveau de la fréquentation de l'école secondaire. Cela peut s'expliquer par la mise en place tardive par le Canada d'une loi sur la fréquentation scolaire minimum basée sur l'obtention du diplôme secondaire. Nous reviendrons sur l'importance de ces variables pour la construction de l'instrument.

¹¹ Selon le recensement de 1996 de Statistiques Canada; 3,197,480 individus de minorités visibles et 4,971,070 immigrants sur une population totale de 29,353,900.

La variable de scolarité de l'individu (EDU, CATEDU) semble indiquer que la loi sur la fréquentation scolaire minimum a porté ses fruits, puisqu'en moyenne les individus ont fréquenté l'école pendant les 12 ans requis pour l'obtention du diplôme secondaire.¹²

Le revenu horaire annuel moyen (REVHORAI), transformé en revenu par année, donne un revenu canadien brut moyen entre 25 000 et 28 000 \$ (le nombre d'heures régulières de travail d'une semaine variant selon la province). Pour une semaine de 40 heures et une année de 48 semaines de travail, cela donne un salaire annuel de 26 700 \$.

Les variables restantes sont des dichotomiques indiquant respectivement le lieu de résidence (HABIATL,...,HABIOUE) et la région où l'individu a effectué sa scolarité (EDUATL,...,EDUOUE). On peut observer que les individus sont bien répartis entre les régions, ce qui nous permettra d'analyser avec plus de justesse nos résultats. Avec ces dernières variables, nous pouvons enfin construire les variables manquantes, indispensables à l'estimation.

3.3. Construction des variables

Pour obtenir la variable d'interaction entre la scolarité du père et la région de résidence, ainsi que pour magnifier l'effet sur la scolarité de l'individu, nous avons dû procéder à quelques manipulations.

Nous avons tout d'abord transformé les catégories d'instruction du père en dichotomique en tassant les deux dernières (deuxième et troisième cycle universitaire) dans la dichotomique

¹² Il faut 11 ans de scolarité au Québec pour l'obtention du diplôme et 12 dans les autres provinces.

diplôme universitaire. Nous avons ensuite vérifié si la région d'instruction correspondait à la région de résidence en procédant à une analyse comparative croisée des variables HABIATL,...,HABIOUE et EDUATL,...EDUOUE.

Le tableau IV reprend les résultats qui nous ont servi dans l'élaboration de la variable instrumentale. On peut voir sur la diagonale le nombre d'individus ayant étudié dans leur lieu de résidence en 1993.

Tableau V : Fréquence croisée entre la région d'instruction et la région de résidence

Fréquence							
%							
% de la ligne							
% de la col.		1	2	3	4		
1		5008	17	161	136	5322	
		21.84	0.07	0.7	0.59	23.21	
		94.1	0.32	3.03	2.56		
		96.31	0.36	2.72	1.92		
2		41	4639	189	80	4949	
		0.18	20.23	0.82	0.35	23.21	
		0.83	93.74	3.82	1.62		
		0.79	97.89	3.19	1.13		
3		114	78	5466	353	6011	
		0.5	0.34	23.84	1.54	26.22	
		1.9	1.3	90.93	5.87		
		2.19	1.65	92.3	5		
4		37	5	106	6498	6646	
		0.16	0.02	0.46	28.34	28.99	
		0.56	0.08	1.59	97.77		
		0.71	0.11	1.79	91.95		
		5200	4739	5922	7067	TOTAUX	

Environ 94% des individus ont étudié dans la région dans laquelle ils habitaient. Cette information est capitale dans l'élaboration de notre instrument. Il aurait été plus simple de prendre comme variable instrumentale l'interaction entre la région d'instruction et la scolarité du

père, mais nous avons besoin de la région d’instruction pour calculer le nombre d’années de scolarité du père.

Le tableau IV nous permet d’affirmer que la région d’instruction correspond aussi à la région de résidence. Nous pouvons donc utiliser la région de résidence dans l’instrument pour mieux tenir compte des effets des différences des systèmes d’éducation que ne le faisaient les catégories fournies dans la base de données de l’EDTR.

Pour amplifier les différences induites par la région d’instruction sur les années de scolarité, nous avons quantifier les catégories EDPERE. . Une analyse croisée semblable à la précédente nous fournit une moyenne d’années de scolarité par catégorie et par région d’instruction. Le tableau 5 met en évidence les différences mentionnées.

Tableau VI : Moyenne d’années de scolarité par région et par catégorie d’instruction

	Edcat1	Edcat2	edcat3	Edcat4	edcat5	Catégorie
Atlantique	6.9255118	10.0864177	11.8179067	12.9487719	18.0511278	
Québec	6.5466948	10.0237778	11.5737952	13.8563183	18.6414474	
Ontario	7.3816054	10.517292	12.1634875	13.9720667	18.5215768	
Prairies	7.252	10.3555316	11.9573084	13.3342063	18.2604938	
Région d’instruction						

Catégorie	Définition
Edcat1:	Primaire
Edcat2:	Secondaire non complété
Edcat3:	Secondaire
Edcat4:	Diplôme professionnel
Edcat5:	Diplôme universitaire

Notons que les variables de type EDCAT sont en fait des dichotomiques qui remplacent la variable EDPERE dont la valeur varie de 1 à 5.

Grâce à ces chiffres nous avons construit une variable de l'éducation du père : EDMOY, qui varie pour chaque individu.

$$\begin{aligned} \text{EDMOY} = & \text{edcat1}(\text{eduat1} * 6.9255118 + \text{eduque} * 6.5466948 + \dots) \\ & + \text{edcat2}(\text{eduat1} * 10.0864177 + \text{eduque} * 10.0237778 + \dots) \\ & + \dots \text{edcat5}(\dots) \end{aligned}$$

La variable EDCAT, qui varie de 1 à 5, est une dichotomique qui définit la catégorie d'instruction du père. Ainsi, on a : $\text{edcat1}=1$ si le père se trouve dans la catégorie 1, $\text{edcat1}=0$ sinon; $\text{edcat2}=1$ si le père se trouve dans la catégorie 2, $\text{edcat2}=0$ sinon, etc...

La dichotomique $\text{edu} + \text{nom}$ de la province définit la valeur de la catégorie d'instruction. Ainsi, $\text{eduque} = 1$ si le père a étudié au Québec, et $= 0$ sinon, $\text{eduat1} = 1$ si le père a étudié dans la région de l'Atlantique, $= 0$ sinon, etc... Cette dichotomique est ensuite multipliée par le nombre d'années de scolarité qui lui correspond. Ce chiffre a été calculé dans le tableau V.

Pour illustrer cette formule, nous prenons l'exemple d'un individu dont le père a un diplôme professionnel de l'Ontario. Dans ce cas, EDMOY va prendre la valeur :

$$\text{EDMOY} = \text{edcat4}(\text{eduont} * 13.9720667) = 13.9720667$$

Nous allons utiliser la variable EDMOY dans l'instrument. La manière dont nous l'avons élaborée, et le fait qu'elle représente bien les différents types d'éducatons des individus, devrait nous permettre d'avoir des résultats très précis.

L'instrument (EDINT) sera simplement composé de EDMOY et des variables de type HABI.

EDINTAB = EDMOY * HABIATL

EDINTQB = EDMOY * HABIQUE

EDINTOB = EDMOY * HABIONT

EDINTPB = EDMOY * HABIOUE

4. Les estimations

4.1. Les résultats

4.1.1. Moindre Carré Ordinaire

Dependent Variable: LOGSAL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	9	1018.75418	113.19491	731.890	0.0001
Error	10566	1634.14814	0.15466		
C Total	10575	2652.90233			
Root MSE		0.39327	R-square	0.3840	
Dep Mean		2.56007	Adj R-sq	0.3835	
C.V.		15.36168			

Tableau VII : Résultats de l'estimation par les MCO

Variable	D.F.	Parameter Estimate	Standard Error	T for Ho: Parameter = 0	Prob > T
Intercept	1	1.182209	0.02609879	45.297	0.0001
Edu	1	0.062667	0.00135552	46.231	0.0001
Sex	1	0.198847	0.00791748	25.115	0.0001
Statmar	1	0.132754	0.00896268	14.812	0.0001
Langmat	1	-0.01745	0.01464151	-1.192	0.2334
Habiatl	1	-0.119732	0.01665253	-7.181	0.0001
Habiont	1	0.032883	0.01693422	1.942	0.0522
Habioue	1	-0.016751	0.01698925	-0.986	0.3242
Exp	1	0.042533	0.00119810	35.501	0.0001
Exp2	1	-0.000695	0.00003023	-22.988	0.0001

4.1.2. Variable Instrumentale

Dependent variable: LOGSAL

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	9	712.32815	79.14757	498.042	0.0001
Error	10566	1679.12141	0.15892		
C Total	10575	2652.90233			
	Root MSE	0.39864	R-Square	0.2979	
	Dep Mean	2.56007	Adj R-SQ	0.2973	
	C.V.	15.57163			

Tableau VIII : Résultats de l'estimation par les VI

Variable	D.F.	Parameter Estimate	Standard Error	T for Ho: Parameter = 0	Prob > T
Intercept	1	0.87068	0.095707	9.097	0.0001
Edu	1	0.085782	0.006962	12.322	0.0001
Sex	1	0.201921	0.008077	25	0.0001
Statmar	1	0.133047	0.009086	14.644	0.0001
Langmat	1	-0.006456	0.015192	-0.425	0.6709
Habiatl	1	-0.100551	0.017824	-5.641	0.0001
Habiont	1	0.025795	0.017293	1.492	0.1358
Habioue	1	-0.011704	0.017286	-0.677	0.4984
Exp	1	0.041521	0.001251	33.196	0.0001
Exp2	1	-0.000636	0.000035208	-18.072	0.0001

Précisons que les estimés ci-dessus sont relatifs aux individus de 16 à 65 ans. Aucune autre restriction n'a été apposée sur le modèle.

Aussi, à des fins comparatives, nous avons estimé les mêmes variables avec les deux méthodes.

4.2. L'analyse des résultats

Le but de cette recherche étant de quantifier le lien causal entre l'éducation et le revenu, nous regarderons ce résultat immédiatement. L'estimation par les Variables Instrumentales de la variable EDUC nous indique un rendement de l'éducation statistiquement significatif d'environ 8,5%. Ce résultat est en soi un peu décevant. On s'attendrait, suite aux dernières recherches (cf. Partie I), à trouver un rendement de l'éducation franchissant la barre des 10% (Card, 1995b et Lemieux et Card, 1998).

L'explication de ce résultat réside dans le choix de l'instrument et dans le choix de la base de données. Card (1998) explique que le choix de l'instrument influence directement les résultats. Dans son étude utilisant la proximité entre une institution universitaire et le lieu de résidence pour instrumenter son estimation, il explique le haut rendement qu'il obtient par le fait que les individus qui habitent proche de l'université ne seraient pas allés étudier s'ils habitaient loin de cette dernière. Les coûts de transports qu'ils ne subissent pas leur rendent accessibles les frais de scolarité. Ainsi, ces individus n'étant pas particulièrement fortunés, le rendement de l'éducation marginal est élevé.

Le même effet peut être observé dans le cas de l'étude de Lemieux et Card (1998) sur le VRA, les militaires ne seraient pas allés à l'école sans le Programme de Réhabilitation des Vétérans (VRA), d'où le haut rendement de l'éducation pour eux.

L'instrument que nous avons choisi ne fait appel à aucun type particulier d'individus, un peu comme l'étude d'Angrist et Krueger (1991a) sur le trimestre de naissance. Comme les individus

impliqués dans cette étude ne présente pas d'avantage ou de désavantage à la fréquentation scolaire, nous pouvons penser que notre estimation du rendement de l'éducation est correcte.

De plus, notre résultat s'approche de celui obtenu par Card (1995a) lorsqu'il modifie son estimation pour tenir compte de la richesse des individus : il obtient alors un rendement de l'éducation de l'ordre de 9,7 % tandis que le nôtre est de 8,5%.

La variable SEX répond bien à nos attentes. L'estimation de 20% indique que les hommes gagnent un salaire de 20% supérieur à celui des femmes. Ce résultat se retrouve dans la quasi-totalité des études sur le revenu.

La valeur positive et significative de l'estimateur de la variable "statut marital" peut être interprétée de différentes façons. Soit les individus mariés sont plus heureux et ainsi sont plus productifs, d'où un salaire plus élevé, soit l'individu avec un salaire plus élevé a plus de chances de se trouver un conjoint (Becker)¹³. Nous pensons que la réalité des choses se trouve entre les deux théories.

Le fait que la variable LANGMAT ne soit pas significative ne nous permet pas d'affirmer quoi que ce soit. Nous avons tenté de prendre des individus plus âgés (43 ans et plus) pour déceler un effet de la langue maternelle, mais les résultats ne sont pas concluants. Nous pensons que cette variable traduirait un effet ségréatif envers les francophones.

¹³ On peut aussi attribuer cette théorie à Darwin : "La pensée de cette lutte universelle provoque de tristes réflexions, mais (...) ce sont les êtres vigoureux, sains et heureux qui survivent et se multiplient".

Les variables de la région de résidence nous indiquent que le fait d'habiter l'est du pays est moins favorable que de résider au Québec. L'interprétation de ces variables se fait en comparaison avec HABIQUE car c'est cette variable qui a été enlevée lors de la régression. On remarque également que le fait d'habiter l'Ontario augmente les chances d'avoir un meilleur salaire. Il faut cependant faire attention avec ces données car l'industrie dans les Maritimes est de nature saisonnière, ce qui aurait tendance à faire diminuer le salaire horaire moyen des individus y habitant. Les autres variables HABIONT et HABIOUE ne sont pas significatives, ce qui laisse croire que les différences avec le Québec ne sont plus très grandes et que la tendance est à l'homogénéisation.

On peut déduire de l'observation des variables EXP et EXP2 (EXP au carré) que l'estimation est réaliste. Nous avons introduit le terme EXP2 pour tenir compte de la courbure de la croissance des salaires en relation avec l'expérience. Les individus avec plus d'expérience gagnent un plus haut salaire, mais l'effet de l'expérience sur le salaire diminue avec l'expérience de l'individu (rendement positif décroissant). Il est donc normal de trouver un terme positif devant EXP et un négatif devant EXP2.

La différence entre l'estimation du rendement de l'éducation par les MCO et celle par les VI ne nous surprend pas. En fin de Partie I, nous avons présenté les explications qui prévalent dans ce genre de problème. Celle qui convient le mieux dans notre cas est celle de Griliches (1977). En effet, l'explication la plus plausible tient au fait que les valeurs fournies par l'individu sur la scolarité de ses parents soient fausses. On retrouve ici les *erreurs de mesures* de Griliches. Ainsi,

le biais positif associé aux erreurs de mesures rends notre estimation par les VI supérieur à celle des MCO.

Les autres explications ne s'appliquent pas pour différentes raisons : pour l'hypothèse de Card parce que le choix de notre instrument n'implique pas une certaine classe d'individus; pour l'hypothèse d'Angrist et Krueger parce que l'effet de la scolarité des parents en interaction avec la province de résidence sur le choix de scolarité de l'individu est non négligeable et pour l'hypothèse de Ashenfelter et Harmon, parce que notre estimation n'est pas particulièrement élevée par rapport aux autres estimations récentes du rendement de l'éducation.

CONCLUSION

Nous avons établis que l'estimation du rendement de l'éducation par les moindres carrés ordinaires était biaisée. Certaines données non observables telles que l'habileté biaisent l'estimation vers le haut. La même estimation effectuée par la méthode des variables instrumentales ne contient pas ce biais, mais peut en contenir d'autres (erreurs de mesures, faiblesse des estimateurs, hétérogénéité de l'instrument). Les plus récentes estimations du rendement de l'éducation par la méthode des variables instrumentales varient entre 5 et 15 %.

Nous avons également montré qu'il était possible d'appliquer cette méthode d'estimation à une base de données canadiennes, nommément l'Enquête sur la Dynamique du Travail et du Revenu, en utilisant les différences dans les systèmes scolaires comme instrument. Nous obtenons ainsi un rendement de l'éducation de 8,5% par la méthode des variables instrumentales et de 6% par les MCO. Le rendement de l'éducation estimé par les MCO est souvent inférieur à celui estimé par la méthode des variables instrumentales. Ce fait peut être observé dans la grande majorité des études fournissant les deux résultats (cf. première partie). Les raisons expliquant cette différence sont nombreuses, mais nous avons retenu celle de Griliches (1977): notre instrument peut être sujet à des erreurs de mesures. Dans la mesure où ce biais vient compenser d'autres biais, notre estimation du rendement de l'éducation est bonne.

Il serait toutefois bon de poursuivre les recherches plus avant afin d'utiliser l'ensemble des données disponibles, c'est-à-dire l'utilisation conjointe des données de 1993 et de 1994 dans la même estimation pour profiter des propriétés des données de Panel.

ANNEXE I: Résultats des régressions donnant l'effet de la scolarité du père sur le niveau d'éducation de l'individu par région

ATLANTIQUE

Dependent Variable: EDU

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	5028.49801	838.08300	111.449	0.0001
Error	3002	22574.75670	7.51991		
C Total	3008	27603.25470			
Root MSE	2.74224	R-square	0.1822		
Dep Mean	11.99408	Adj R-sq	0.1805		
C.V.	22.86331				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	9.630029	0.17027224	56.557	0.0001
SEX	1	-0.209470	0.10591491	-1.978	0.0481
STATMAR	1	0.253713	0.12340468	2.056	0.0399
LANGMAT	1	0.015985	0.15383897	0.104	0.9173
EXP	1	0.087677	0.01501431	5.840	0.0001
EXP2	1	-0.002781	0.00036192	-7.685	0.0001
EDPERE	1	1.079591	0.04702998	22.955	0.0001

QUÉBEC

Dependent Variable: EDU

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	6884.36384	1147.39397	110.229	0.0001
Error	2883	30009.80016	10.40923		
C Total	2889	36894.16400			
Root MSE	3.22633	R-square	0.1866		
Dep Mean	12.13294	Adj R-sq	0.1849		
C.V.	26.59152				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	10.342612	0.32693646	31.635	0.0001
SEX	1	0.185756	0.12712492	1.461	0.1441
STATMAR	1	0.045430	0.13956801	0.326	0.7448
LANGMAT	1	-0.091226	0.26020187	-0.351	0.7259
EXP	1	0.099152	0.01787588	5.547	0.0001
EXP2	1	-0.003963	0.00042030	-9.429	0.0001
EDPERE	1	0.996677	0.05407029	18.433	0.0001

ONTARIO

Dependent Variable: EDU

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	4619.78933	769.96489	109.295	0.0001
Error	2980	20993.54063	7.04481		
C Total	2986	25613.32996			
Root MSE	2.65421	R-square	0.1804		
Dep Mean	13.19451	Adj R-sq	0.1787		
C.V.	20.11599				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	11.306224	0.17357010	65.139	0.0001
SEX	1	0.301198	0.10099250	2.982	0.0029
STATMAR	1	0.390069	0.11420863	3.415	0.0006
LANGMAT	1	-0.330384	0.16804826	-1.966	0.0494
EXP	1	0.030142	0.01480857	2.035	0.0419
EXP2	1	-0.001810	0.00034193	-5.293	0.0001
EDPERE	1	0.807886	0.04072790	19.836	0.0001

OUEST

Dependent Variable: EDU

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	3594.30336	599.05056	103.399	0.0001
Error	3952	22896.26506	5.79359		
C Total	3958	26490.56842			
Root MSE	2.40699	R-square	0.1357		
Dep Mean	12.90063	Adj R-sq	0.1344		
C.V.	18.65791				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	11.035148	0.13244604	83.318	0.0001
SEX	1	-0.118420	0.08070279	-1.467	0.1424
STATMAR	1	0.076265	0.08956784	0.851	0.3946
LANGMAT	1	0.009366	0.19497322	0.048	0.9617
EXP	1	0.085132	0.01172130	7.263	0.0001
EXP2	1	-0.002540	0.00027950	-9.087	0.0001
EDPERE	1	0.671471	0.03136912	21.405	0.0001

ANNEXE II : Équivalence entre les variables de l'EDTR et les nôtres ainsi que leur significations

EDTR ESTIMATION DEFINITION

eage26b	Age	Age de l'individu en années
sex21	Sex	Sexe – transformé en dichotomique
marst26b	Statmar	État Matrimonial - transformé en dichotomique
motn2g15	Langmat	Langue Maternelle - transformé en dichotomique
vismn15	Minvis	Statut de Minorité visible selon le gouvernement
edfath21	Edpere	Éducation du père en années
edmoth21	Edmere	Éducation de la mère en années
immst15	Immigra	Statut d'immigrant
regre25b	Provhabi	Région de résidence
cmphw28b	Revhorai	Taux horaire de rémunération annuelle (moy.)
yrxft11b	Exp	Expérience en années
studf26b	Dicoedu	Dichotomique indiquant si l'individu étudie
pvelh18b	Provedu	Province d'éducation
yrsch18b	Edu	Éducation en années
hlevg18b	Catedu	Éducation en catégories

BIBLIOGRAPHIE

ALTONJI, Joseph G. et Thomas A. DUNN (1996), “ The Effect of Family Characteristics on the Return to Schooling ”, *Review of Economics and Statistics* 78 : 692-704.

ANGRIST, Joshua D., et Alan B. KRUEGER (1991), “ Does Compulsory Schooling Affect Schooling and Earnings? ”, *Quartely Journal of Economics* 106 : 979-1014

ANGRIST, Joshua D., et Alan B. KRUEGER (1992), “ Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery ”, National Bureau of Economic Research Working Paper No.4067 , Août.

ANGRIST, Joshua D. et Guido W. IMBENS (1995), “ Two Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity ”, *Journal of the American Statistical Association* 90 : 431-442.

ANGRIST, Joshua D., Guido W. IMBENS et Donald B. RUBIN (1996), “ Identification of Causal Effects Using Instrumental Variables ”, *Journal of the American Statistical Association* 91 : 444-472.

ASHENFELTER, Orley et Cecilia E. ROUSE (1998), “ Income, Schooling and Ability: Evidence from a New Sample of Identical Twins ”, National Bureau of Economic Research no.6106, Juillet.

ASHENFELTER, Orley et Cecilia ROUSE (1999), “ Schooling, Intelligence, and Income in America: Craks in the Bell Curve ”, National Bureau of Economic Research Working Paper No.6902, Janvier 1999.

BECKER, Gary S. (1964), Human Capital : A Theoretical and Empirical Analysis, with Special Reference to Education, Columbia University Press, New York, 1964.

BOUND, John, David A. JAEGER et Regina M. BAKER (1995), “ Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak ”, *Journal of the American Statistical Association* 90 : 443-450.

CARD, David (1995a), “ Earning, Schooling and Ability Revisited ”, *Research in Labor Economics* Vol 14, JAI Press, Solomon Polachek, Greenwich Connecticut, 1995, pages 23-48.

CARD, David (1995b), “ Using Geographic Variation in College Proximity to Estimate the Return to Schooling ”, *Aspects of Labour Market Behavior: Essays in Honor of John Vanderkamp*, University of Toronto Press, Louis N. Christofides, E. Kenneth Grant, et Robert Swidinsky, Toronto, 1995, pages 201-222.

CARD, David (1998), “ The Causal Effect of Education on Earnings ”, *Handbooks of Labor Economics*, North Holland, Amsterdam et New York, Orley Ashenfelter et David Card, à paraître.

DARWIN, Charles (1859), “ L’origine des espèces au moyen de la sélection naturelle ou la préservation des races favorisées dans la lutte pour la vie ”, GF-Flammarion, Paris, 1992, traduction d’Edmond Barbier.

GOLDWIN, Claudia et Lawrence KATZ (1997), “ Why the United States Led in Education: Lessons from Secondary School Expansion, 1910 to 1940 ”, National Bureau of Economic Research Working Paper No.6144, Août.

GRILICHES, Zvi (1977), “ Estimating the Return to Schooling : Some Econometric Problems ”, *Econometrica* 45 : 1-22.

GRILICHES, Zvi (1979), “Sibling Models and Data in Economics : Beginning of a Survey”, *Journal of Political Economy* 87 : S37-S65.

HECKMAN, James J. et Solomon POLACHEK (1974), “Empirical Evidence on the Functional Form of the Earning-Schooling Relationship”, *Journal of the American Statistical Association* 69 : 350-354.

LEMIEUX, Thomas et David CARD (1998), “Education, Earnings and the Canadian G.I.Bill”, National Bureau of Economic Research Working Paper No. 6718, Septembre 1998.

MINCER, Jacob (1974), *Schooling, Experience and Earnings*, Columbia University Press, New York, 1974.

MURPHY, Kevin M. et Finis WELCH (1990), “Empirical Age-Earnings Profiles”, *Journal of Labor Economics* 8 : 202-229.

STEINER, George, “Errata, Récit d’une Pensée”, Gallimard, Paris, 1998, page 66.

TOYNBEE, Arnold, “La Grande Aventure de l’Humanité”, Éditions Payot et Rivages, Paris, 1994, p.12

