

Um Estudo Sobre Retorno em Escolaridade no Brasil*

Adolfo Sachsida**

Paulo Roberto Amorim Loureiro***

Mário Jorge Cardoso de Mendonça****

Sumário: 1. Introdução; 2. Metodologia; 3. Base de dados; 4. Resultados econométricos; 5. Conclusões.

Palavras-chave: retorno em escolaridade; vies de seleção amostral; pseudo painel; vies de habilidade; equação de salários.

Códigos JEL: C31; I29.

O objetivo deste trabalho é investigar o retorno em escolaridade para o Brasil. Para tal, verifica-se por diferentes procedimentos se diferentes fontes de vies estariam prejudicando a estimação fidedigna para a equação de salários, e com isso gerando vies para o retorno em escolaridade. O primeiro método tem por base o estimador de Heckman (1979) sugerindo que o vies possa ser causado pela estratégia de “job search” que faz com que o salário não dependa apenas da oferta de trabalho como também da procura por emprego. A segunda abordagem faz uso da metodologia de Garen (1984) que permite tratar o grau de escolaridade como uma escolha racional do agente. Por fim, investiga-se a existência de vies de variável omitida causado pela ausência de uma variável que possa medir a habilidade do indivíduo (Griliches, 1977). O tratamento empírico neste caso foi feito com base na metodologia de pseudo painel (Deaton, 1985). Cada um dos resultados gerados por esses três estimados são comparados com os obtidos por OLS. Entre as principais conclusões desta pesquisa pode-se destacar que a evidência acerca da hipótese de vantagens comparativas de Willis e Rosen (1979), indícios de endogeneidade na escolha da escolaridade e pouca importância para o vies de variável omitida.

The objective of this paper is to investigate the returns to schooling issue for Brazil based on distinct methodologies in order to reach a reliable estimate for this type of measurement. The first one suggests that bias is caused by the job search strategy adopted

* Artigo recebido em abr. 2003 e aprovado em fev. 2004.

** Universidade Católica de Brasília. Mestrado em Economia de Empresas.
E-mail: sachside@pos.ucb.br

*** Universidade Católica de Brasília. Mestrado em Economia de Empresas.

**** Instituto de Pesquisa Econômica Aplicada. E-mail: mjorge@ipea.gov.br

by the agent. In this case, wages do not depend solely on the wage supply but also on the job-search strategy (Gronau, 1974). Second, we use the methodology developed by Garen (1984) that allows dealing with the optimal choice of years of schooling due to the fact that this variable is continuous and ordered. Lastly, a pseudo-panel analysis (Deaton, 1985) is conducted in an attempt to deal with the relevant variable omission related to “ability bias” (Griliches, 1977) in the model. All results are compared to returns to schooling equation generated by OLS. Among main conclusions we have evidences about comparative advanced from Willis e Rosen (1979), endogeneity on schooling’ decision and no relevant evidence on the existence of “ability bias”.

1. Introdução

Nos últimos anos tem se verificado um grande interesse no estudo sobre a relação entre escolaridade e o sucesso no mercado de trabalho (Card, 2001). A origem desse interesse pode estar baseada na busca por uma explicação do aumento nos diferenciais de salários entre pessoas com alto e baixo nível educacional, o que poderia indicar retornos diferentes entre grupos educacionais distintos (Katz e Autor, 1999). Outra explicação para o crescente interesse sobre esse tema se deve a estudos que mostram a importância do capital humano no desenvolvimento de um país (Topel, 1999).

Dos temas relacionados à questão da escolaridade, o retorno em escolaridade tem recebido destacada atenção por parte dos economistas. As pesquisas neste campo remontam aos trabalhos de Becker (1962, 1975) e Mincer (1958, 1974). Becker (1975). apresenta um modelo onde os agentes decidem investir em capital humano tendo em vista o custo e o benefício de cada ano a mais de escolaridade. A partir disso toda uma linha de pesquisa começou a se desenvolver, tendo como ponto principal o tratamento econométrico do problema. Como assinala Griliches (1977), grande parte do trabalho aplicado se concentra em estimar uma versão da seguinte equação:

$$w_i = \ln W_i = \alpha + \beta S_i + \delta X_i + u_i \quad (1)$$

onde:

W é uma medida de renda ou salário;

S é uma medida de escolaridade tomada em geral em número de anos de escolaridade ou graus completados;

X um conjunto de variáveis de controle que podem ter efeito sobre a renda, e u

é o distúrbio aleatório que comporta todas as forças não diretamente explicitadas no modelo, mas que possuem influência sobre os ganhos do indivíduo. Costuma-se denominar a equação (1) como equação de salários ou equação minceriana. Nela, o retorno em escolaridade é definido pelo parâmetro β , que representa a variação marginal da renda em relação à escolaridade. No entanto, como o próprio Griliches (1977) coloca, existem muitos pontos específicos que podem ser levantados no sentido de questionar os fundamentos sobre os quais este modelo se assenta.

Os desenvolvimentos posteriores da econometria possibilitaram avanços significativos no tratamento dos problemas econométricos relacionados à equação (1). Tal fato foi também um dos determinantes do ressurgimento do interesse neste tema (Card, 2001). Grande parte da literatura passou a considerar a hipótese proposta por Becker (1975) de que a decisão de estudo como endógena (Willis e Rosen, 1979, Garen, 1984, Kenny et alii, 1979). Além disso, houve outros avanços no sentido de verificar se a estimação da equação (1) poderia estar viesada ainda devido a outros fatores. Heckman (1974) testa a hipótese de que o agente tenha implícito um determinado salário de reserva, abaixo do qual ele não aceitaria participar do mercado de trabalho. Alguns autores assinalam ainda que existe um tipo de problema relacionado ao chamado “viés de habilidade”. Esse viés decorre da omissão de uma variável relevante no modelo, que possa traduzir a habilidade ou talento do indivíduo (Griliches, 1977, Garen, 1984, Hausman e Taylor, 1981).

O objetivo deste trabalho é investigar a questão do retorno em escolaridade para o Brasil. A proposta fundamental aqui é verificar quais as fontes de viés que poderiam estar afetando a estimação fidedigna da equação (1). Neste sentido, a estratégia usada neste trabalho é elaborar uma análise comparativa entre os resultados obtidos a partir de métodos empregados para correção de viés específico com aqueles obtidos por OLS. As fontes de viés estudadas nesta pesquisa são a ocorrência de salário de reserva no mercado de trabalho, existência de endogeneidade na decisão por estudo e, por fim, verificar a existência na equação (1) do chamado “viés de habilidade”.

Este artigo está estruturado da seguinte maneira. Além dessa introdução, a seção dois tem como objetivo aprofundar a discussão referente aos diferentes métodos empregados para correção dos diferentes casos de endogenia que podem estar presentes na equação de salários. A seção três apresenta a base de dados que dá suporte empírico a este trabalho. A seção quatro apresenta e discute os resultados econométricos obtidos a partir da aplicação das diversas metodologias utilizadas. Por fim, na seção cinco, são apresentadas as principais conclusões do trabalho.

2. Metodologia

A primeira fonte de viés sugere que o mesmo seja causado pela estratégia de demanda por emprego do agente. Nesse caso, o salário não depende apenas da oferta de salário, mas também da estratégia de “job-search” do agente (Gronau, 1974). Assim, como assinala Heckman (1974) a idéia que aparece é que na sua estratégia de demanda por emprego, o agente tem implícito um determinado salário de reserva abaixo do qual ele não aceitaria participar do mercado de trabalho. Dessa maneira, a não consideração desse fato faz com que as estimativas obtidas por OLS para a equação (1) sejam viesadas devido ao aparecimento daquilo que se denomina de viés de seleção amostral, pois grupos que usam regras distintas são analisados de modo idêntico. Com intuito de observar a existência deste problema, faz-se uso da metodologia desenvolvida por Heckman (1979).

A segunda fonte de viés a ser objeto de investigação deste estudo segue a linha original de Becker (1975) no sentido de verificar se existe endogenia quanto à escolha do nível de escolaridade. Diferentemente do tratamento empreendido por Willis e Rosen (1979), onde escolaridade foi enquadrada em duas categorias estanques, esse trabalho faz uso da metodologia desenvolvida por Garen (1984), que permite tratar a escolha ótima dos anos de escolaridade tomando como base o fato desta variável ser contínua e ordenada, o que faz com que o problema seja tratado com grau bem maior de realismo.

Por fim, a última fonte de viés estudada deriva do problema relacionado ao chamado “viés de habilidade”. Esse viés decorre da omissão de uma variável relevante no modelo representado pela equação (1) que possa representar a habilidade ou talento do indivíduo (como sua inteligência, por exemplo). A omissão desta variável pode se dever a vários fatores. Primeiro, não existe consenso quanto ao modo de mensuração desta variável. Alguns estudos tentam contornar este problema construindo medidas de habilidade (Heckman et alii, 2000) e Tobias (2003) ou então usando o QI como *proxy* desta variável (Garen, 1984) e (Griliches, 1977). Deve ser mencionado que é comum na literatura sobre retorno em educação, o uso de variáveis instrumentais para corrigir o viés de variável omitida. No entanto, como assinala Card (2001) a principal crítica que recai sobre esse método deriva da escolha adequada dos instrumentos que nem sempre é trivial.

Tendo em vista ser a habilidade uma característica própria do indivíduo e, por hipótese, não se alterar no tempo, uma maneira apropriada de corrigir esse tipo de viés se dá a partir do emprego da metodologia de dados em painel, tal como aparece em Hausman e Taylor (1981). Infelizmente, no Brasil, não existe uma pesquisa que faça um acompanhamento temporal do mesmo indivíduo, o que impede uma

aplicação direta dessa técnica. Uma maneira alternativa de contornar a falta de uma amostra específica se dá através da aplicação da técnica de pseudo painel (Deaton, 1985). Desta forma, os resultados estimados para a equação (1) a partir deste método também são apresentados neste estudo. Assim, foram estimados quatro modelos distintos para verificar a robustez dos resultados. Além disso, para verificar a sensibilidade do resultado devido à inclusão de variáveis, cada modelo foi estimado com diferentes conjuntos de variáveis explicativas. Os modelos podem ser enquadrados em quatro categorias distintas quanto à metodologia de estimação: i) OLS; ii) Heckman (1979); iii) Garen (1984) e, iv) pseudo painel.

Em relação aos estudos empreendidos para o Brasil podem ser feitas algumas citações. Ueda e Hoffmann (2002) usando o método de variáveis instrumentais encontram uma taxa de retorno da escolaridade em torno de 9,8%. Fazendo uso do procedimento de Heckman (1979), Loureiro e Galvão (2001) estimam o retorno em escolaridade para os setores rurais e urbanos. Nesse estudo o retorno em escolaridade para homens no setor urbano foi de 18,58%, enquanto que no setor rural o retorno se situou em 11,35%. Silva e Kassouf (2000), investigaram a existência de retornos diferenciados (por sexo) à escolaridade e treinamento. Os resultados evidenciam que a taxa de retorno da escolaridade relativo às mulheres foi maior que a dos homens. Em um estudo sobre a existência de dualidade no mercado de trabalho brasileiro, Soares e Gonzaga (1999), utilizando uma metodologia proposta por Dickens e Lang (1985, 1992), estimaram equações de salários a partir de um modelo de “*switching-regressions*”. As taxas de retorno estimadas para escolaridade variaram de 12% (setor secundário) a 23,4% (setor primário) ao ano. Enquanto, a taxa de retorno da educação pelo modelo competitivo (uni-setorial) linear foi de 15,4%.

3. Base de Dados

Os dados utilizados neste estudo foram obtidos da Pesquisa Nacional por Amostra de Domicílio (PNAD). Excetuando-se obviamente os resultados gerados por pseudo painel, todas as demais estimações foram obtidas a partir de duas amostra distintas: cross section e dados empilhados (pooling) de 1992 a 1999. Aqui os dados de salários foram deflacionados em relação ao ano base de 1996 pelo IGP. Para os dados em *cross-section* o estudo tomou por base as informações referentes ao ano de 1996.¹ O uso da amostra expandida de dados empilhados tem como

¹A escolha desse ano se deve ao fato dele estar no ponto mediano do período da estimativa com dados empilhados.

objetivo observar se tal procedimento incorpora algum tipo de informação nova ao modelo.

Uma etapa importante para a correta estimativa dos retornos em escolaridade, consiste na seleção da amostra. Apesar de estarmos trabalhando com dados da PNAD, nem todas as observações podem ser usadas. Dessa maneira, alguns filtros devem ser realizados. Em primeiro lugar, nossa amostra é composta apenas por homens entre 24 e 56 anos de idade, pois suas decisões, sobre o nível de escolaridade, são menos complicadas por considerações de fertilidade (Cameron e Heckman, 2001).² O segundo filtro implica que, nessa amostra, serão incluídas apenas pessoas que não estejam estudando.³ O terceiro filtro age da seguinte maneira, quando uma observação não possui uma, ou mais, informações sobre as variáveis independentes, então essa observação é retirada da amostra. Esse procedimento aparece em Heckman et alii (2000) e Soares e Gonzaga (1999).⁴ O quarto filtro tenta evitar que sejam incluídas na amostra pessoas que: i) possuam um salário extremamente alto que possa viesar os resultados; e ii) que não estejam trabalhando.⁵ Dessa maneira, esse filtro faz com que a amostra seja composta apenas de indivíduos que recebam um salário horário entre R\$ 1,00 e R\$ 500,00.⁶ Além disso, devido à dinâmica própria que rege a administração pública e o setor agrícola da economia brasileira, as pessoas ocupadas nestes setores foram também excluídas da amostra (Soares e Gonzaga, 1999).

Conforme foi salientado na seção anterior, no Brasil não existe uma pesquisa que faça acompanhamento do mesmo indivíduo ao longo do tempo o que impossibilita a aplicação direta da técnica de dados em painel que seria a mais adequada para correção de vies de variável omitida relativa à habilidade do indivíduo (Hausman e Taylor, 1981). A alternativa encontrada foi a de montar um *pseudo painel* seguindo a metodologia que aparece em Deaton (1985). O período de abrangência da amostra foi de 1992 a 1999 (com exceção de 1994, quando a PNAD não foi realizada). A montagem do *pseudo painel* se deu da seguinte forma. Foram feitos três cortes na amostra, que consistiu em dividir a amostra para cada ano em três

²Bratsberg e Terrell (2002), Heckman et alii (2000), Soares e Gonzaga (1999) e Garen (1984), também restringem sua amostra a apenas homens.

³Bratsberg e Terrell (2002), Heckman et alii (2000) e Garen (1984) também utilizam esse procedimento.

⁴Garen (1984) estima duas regressões: a primeira usando observações incompletas e a segunda fazendo uso de uma amostra que considera apenas as observações com todas as informações disponíveis.

⁵Para possibilitar a estimação do modelo de Heckman (1979) foi necessário permitir que homens que não trabalham fizessem parte da amostra.

⁶Heckman et alii (2000) restringem sua amostra a pessoas que recebem um salário horário entre U\$ 1,00 e U\$ 100,00.

grupos classificados da seguinte maneira: Grupo 1: pessoas entre 24 e 34 anos de idade; Grupo 2: com idade entre 35 e 45 anos; e Grupo 3: pessoas entre 46 e 56 anos de idade. Uma vez elaborado isso, a tarefa a seguir diz respeito aplicar as técnicas específicas de dados em painel para essa base de dados que foi montada.

De modo a obter estimativas consistentes para o modelo. A abordagem clássica de dados em painel trata de verificar se o componente individual omitido é ou não correlacionado com algum regressor. No primeiro caso, o modelo deve ser estimado por meio da aplicação de um estimador denominado de efeito fixo. No segundo caso, o correto o mais apropriado é estimar o modelo por efeito aleatório. Para verificar qual das duas hipóteses é a que melhor se aplica, emprega-se o teste de Hausman (Hausman e Taylor, 1981). No caso em questão, onde a habilidade é o componente omitido próprio do indivíduo, é razoável a colocação de que ele seja correlacionado com a escolaridade. Uma hipótese razoável é que indivíduos com maior habilidade optam por estudar mais de modo a exercer sua vantagem comparativa.

4. Resultados Econométricos

Nesta seção são apresentados os resultados econométricos obtidos com base nas metodologias propostas neste estudo. Adiante, na tabela 3, são mostrados todos os resultados para a equação de salários. No entanto, inicialmente serão apresentados os resultados intermediários referentes à equação para a determinação dos anos de escolaridade para o modelo de Garen (1984) e a equação de seleção do modelo de Heckman (1979). A equação para escolaridade foi estimada com base na proposta original de Garen (1984)⁷ tal que,

$$S = \alpha + \beta_1 \text{Raça} + \beta_2 \text{Filhos} + \beta_3 (SM) + \eta \quad (2)$$

onde:

S representa os anos de escolaridade;

β_1 é uma constante,

Raça é uma *dummy* que assume valor 1 se o respondente é branco e valor 0 caso contrário, Filhos é uma *dummy* que assume valor 1 se o respondente tem filhos e valor 0 caso contrário,

SM representa a escolaridade da mãe e

⁷Aqui por economia foram omitidos os detalhes técnicos acerca do método de Garen e que pode ser obtido no trabalho original deste autor.

η representa o distúrbio aleatório. Dando ainda prosseguimento a metodologia de Garen (1984), a equação de salários foi estimada da seguinte maneira:

$$\begin{aligned} \log w &= \beta_0 + \beta_1 S + \beta_2 Exp + \beta_3 (Exp)^2 + \beta_4 Raça + \beta_5 (S \times Raça) \\ &+ \beta_6 (S \times Exp) + \beta_7 (Exp \times Raça) + \beta_8 (S \times Raça \times Exp) \\ &+ \beta_9 \hat{\eta} + \beta_{10} (\hat{\eta} \times S) + \epsilon \end{aligned} \quad (3)$$

onde:

$\log w$ é o logaritmo do salário mensal,

β_0 é uma constante,

Exp são os anos de experiência do trabalhador,⁸

$\hat{\eta}$ é o resíduo estimado da equação de escolaridade. Além disso, foram inclusos alguns termos de interação na equação.

ϵ representa o erro da estimativa.

Para verificar a robustez das equações acima foram inclusas algumas variáveis, que não foram objeto de apreciação no estudo de Garen (1984) e que, no entanto, podem ter influência sobre a decisão sobre escolaridade. Dessa maneira, a equação para escolaridade foi acrescida das variáveis dummy para estado civil (csdo) que assume valor 1 se o indivíduo é casado e 0 caso contrário,⁹ e das dummies para regiões (a região sudeste foi adotada como base) de modo a captar a grande heterogeneidade existente no mercado de trabalho brasileiro. Já em relação à equação de salários existiu o acréscimo das seguintes, variáveis, uma dummy para setor formal (formal) que assume valor 1 se o trabalhador trabalha neste setor e 0 caso contrário, e outra dummy para sindicato (sind) que assume valor 1 se o trabalhador é filiado a um sindicato e 0 caso contrário. Os resultados para a equação de escolaridade são mostrados na tabela 1.

Conforme pode ser observado na tabela 1, não existe diferença significativa para as duas amostras. Ambas as regressões apresentam o mesmo resultado para a escolaridade da mãe. Tal como era esperado, os resultados indicam que o nível de escolaridade da mãe afeta positivamente a escolaridade do indivíduo. Verifica-se igualmente um efeito positivo para a variável raça, sinalizando que pessoas brancas possuem um nível de escolaridade superior aos não brancos. Já para a variável filhos, é de se esperar que afete negativamente a escolaridade do indivíduo. Todos estes resultados estão em concordância com aqueles que aparecem em Garen

⁸Seguindo o trabalho de Heckman et alii (2000) essa variável foi calculada da seguinte maneira: $Exp = Idade - escolaridade - 6$.

⁹Infelizmente não há como saber se o indivíduo estava ou não estudando quando se casou.

(1984). No que se refere a variável *csdo*, o sinal negativo para o seu coeficiente mostra que o casamento que pode interferir negativamente na decisão por escolaridade. Este fato está em acordo com o que se verifica nas sociedades mais desenvolvidas onde as pessoas postergam a vida conjugal e também a fecundidade para poderem alcançar um maior nível de bem-estar. Por fim, o fato das dummies regionais apresentarem significância assinala a heterogeneidade existente entre elas.

Tabela 1
Equação para Escolaridade para o Modelo de Garen[#]

Variáveis Independentes	Garen Cross-Section 1996	Garen Empilhados 1992-99
Constante	15,289	15,023
Raça	1,376	1,479
Filhos	-1,153	-0,998
SM	0,207	0,209
Csdo	-0,396	-0,147
Norte	-0,389	-0,192
Nordeste	-0,477	-0,392
Sul	-0,712	-0,642
Centro	-0,763	-0,533
R² ajustado	0,216	0,216
Observações	15.704	111.074

[#]: Todas as variáveis presentes foram significantes a um nível de 1% de significância

Heckman (1979) assinala que apesar da estimativa por OLS poder oferecer informações interessantes acerca dos determinantes da renda do indivíduo, existe a possibilidade de que o coeficiente relativo ao retorno em escolaridade possa ter sido estimado de modo viesado. Um motivo estaria no fato do salário depender não apenas da oferta de emprego como também da estratégia de “job-search” do agente (Gronau, 1974). A idéia aqui é que o agente tem implícito um salário de reserva, abaixo do qual o agente não aceita participar do mercado de trabalho. Desta maneira, a escolha de participar do mercado de trabalho seja regida por uma regra de decisão que faz com que o agente aceite ou não o salário ofertado. Se esse for o caso, a amostra estaria poluída pelo chamado viés amostral. Heckman (1979) desenvolve uma metodologia própria para tratar esse tema, na qual a equação de salário aparece associada à equação de seleção que por sua vez, define a decisão de participar do mercado de trabalho.

A tabela 2 apresenta os resultados obtidos para a equação de seleção para o modelo de Heckman (1979). De acordo com os resultados desta tabela o teste de razão de máxima verossimilhança para a estatística rho, que mede a correlação entre a equação de seleção e da determinação do salário, indica que existe viés de

seleção de amostra no modelo. Conforme se pode ainda notar, alguns resultados se encontram em concordância com os observados na literatura. A escolaridade tem efeito positivo sobre o salário de reserva do agente, variável latente da equação de seleção. Diferentemente, a variável filhos tem efeito negativo sobre o salário de reserva, resultado já observado em Kassouf (1994), Silva e Kassouf (2000) e Loureiro e Galvão (2001) em estudos feitos para o Brasil. Este resultado é intuitivo na medida que ter filhos leva a uma maior exigência para que ele se incorpore a força de trabalho e, portanto isto faz com que o custo de oportunidade de ser manter fora do mercado de trabalho seja maior.

Ainda, na tabela 2, pode-se perceber que existe ganho de informação advindo do uso da amostra com dados empilhados (pooling). Isso se faz mais presente quando se observa a significância do coeficiente da variável raça que é significativa a partir dos resultados obtidos por pooling, mas não encontra significância para dados de cross-section. Aqui o sinal encontrado está em conformidade com o esperado. O sinal positivo para a raça mostra que o fato do indivíduo ser de cor branca tem efeito positivo sobre seu salário de reserva.

Por fim, a dificuldade está em explicar o efeito negativo que a escolaridade da mãe e o fato de ser casado têm sobre o salário de reserva que a princípio pode parecer contra factual ou pouco intuitivo, visto ainda que Silva e Kassouf (2000) obtêm sinal positivo para a variável csdo. No entanto, pode não ser esse o caso. Com relação à escolaridade da mãe, os estudos conhecidos feitos para o Brasil para explicar a participação na força de trabalho não levam em consideração essa variável (Kassouf, 1994, Silva e Kassouf, 2000, Loureiro e Galvão, 2001) e, portanto não se pode dizer nada de conclusivo a priori. Para a variável csdo, é possível que ser casado aumente o custo de oportunidade de ser manter fora do emprego, o que é um resultado intuitivo pela maior exigência por emprego advinda das necessidades da vida conjugal, principalmente para o homem.

A tarefa a seguir diz respeito à análise dos resultados para o coeficiente do retorno em escolaridade obtido a partir dos diversos modelos. Tais resultados estão expostos na tabela 3. As estimativas com base nos métodos OLS (modelos 1 e 2), Heckman (3 e 4) e Garen (5 e 6) foram estimados tanto para dados de cross-section, para o ano de 1996, como para dados empilhados, no período 1992-1999. O modelo de pseudo painel (7) apresenta o resultado apenas para o estimador de efeitos fixos.¹⁰ Naturalmente, as estimativas por OLS são reportadas para servirem de base de comparação com os outros métodos de estimação.

¹⁰O teste de Hausman rejeita a hipótese nula a um nível de 1% de significância de que o componente individual seja não correlacionado com os regressores. Isto significa que os resultados gerados pelo estimador de efeito aleatório são viesados.

Tabela 2
Equação de Seleção para o Modelo de Heckman[#]

Variáveis Independentes	Heckman Cross-Section (1996)	Heckman Pooling 1992-1999
Constante	2,373	2,031
S	0,021	0,020
Raça	0,016*	0,074
Filhos	-0,630	-0,464
SM	-0,013	-0,005
Csdo	-0,154*	-0,073
Norte	-0,080*	0,051*
Nordeste	0,034*	0,044
Sul	-0,049*	-0,055
Centro	-0,098*	0,013*
ρ	0,583	0,542
Teste Razão Max.	Ho: $\rho = 0$	Ho: $\rho = 0$
Verossimilhança	Chi ² (1) = 61, 29 Pr > Chi ² = 0,000	Chi ² (1) = 60, 07 Pr > Chi ² = 0,000
Tamanho da Amostra	40070	286637

[#] Variável latente: salário de reserva. * Variáveis estatisticamente não significativas ao nível de 5%. A exclusão de tais variáveis pouco afeta os resultados apresentados. As demais variáveis são todas significativas ao nível de 5%.

Um exame mais detalhado da tabela 3 nos informa sobre a robustez dos resultados. Note que, apesar do grande número de modelos e métodos empregados, os coeficientes estimados são quase sempre muito próximos e, em via de regra, apresentam o mesmo sinal. Além disso, tal como aparece na literatura, os coeficientes das variáveis escolaridade e experiência são sempre positivos, enquanto o coeficiente da experiência ao quadrado é sempre negativo. Ainda na tabela 3, podemos observar, tal como se observa em Garen (1984) que o coeficiente de $\hat{\eta}$ é sempre negativo e significativo, ao passo que o efeito de $\hat{\eta}$ é sempre positivo e significativo. Isto é consistente com a hipótese de vantagens comparativas proposta por Willis e Rosen (1979). Resíduos altos e positivos na equação de escolaridade ($\hat{\eta}$'s elevados) tendem a estar associados com menores vencimentos para níveis menores de escolaridade. Observa-se ainda que este efeito diminui à medida que o nível de escolaridade aumenta. Isto indica que aqueles indivíduos com níveis inesperadamente grandes de escolaridade (altos $\hat{\eta}$'s), teriam recebido menos que os outros se tivessem adquirido menos escolaridade. Entretanto, para níveis altos de escolaridade, eles tendem a receber mais que os outros receberiam. Sendo que o oposto ocorre para baixos valores de $\hat{\eta}$'s (Garen, 1984).

Outras informações interessantes, presentes na tabela 3, mostram o efeito positivo do casamento, de pertencer ao setor formal e da filiação a sindicatos sobre o salário. Note que em todas as equações em que aparecem, essas variáveis são

positivas e significantes. Cabe ressaltar que, apesar da similaridade com outros trabalhos a respeito dos sinais das variáveis, os coeficientes desse estudo não são diretamente comparáveis com outros. Isto ocorre devido à inclusão de alguns termos de interação nas regressões. Tais termos afetam a magnitude e a significância dos parâmetros.

Tabela 3
Equação para o logaritmo do salário mensal

Modelo	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Metodologia:	OLS	OLS	Heckman	Heckman	Garen	Garen	Ef. Fixo
Dados:	C-S	Pooling	C-S	Pooling	C-S	Pooling	Pooling
Período	1996	1992-99	1996	1996	1996	1992-99	1992-99
Constante	4,743	4,579	4,508	3,842	4,703	4,690	—
S	0,141	0,144	0,129	0,147	0,139	0,130	0,160
Exp	0,063	0,069	0,070	0,085	0,048	0,048	0,094
Exp ²	-0,0007	-0,0008	-0,0009	-0,001	-0,0006	-0,0006	-0,001
Raça	-0,341	-0,321	-0,248	-0,175	-0,369	-0,338	-0,309
SxRaça	0,046	0,045	0,047	0,040	0,047	0,047	0,044
SxExp	-0,001	-0,001	-0,001	-0,001	-0,0003*	-0,0002	-0,002
ExpxRaça	0,010	0,007	0,008	0,006	0,013	0,011	0,007
SxRaçaxExp	-0,0002*	-0,0002*	-0,0006	-0,0004	-0,0009	-0,0009	-0,0002*
η	—	—	—	—	-0,110	-0,104	—
$\eta \times S$	—	—	—	—	0,009	0,009	—
Csdo	0,197	0,169	0,324	0,396	0,194	0,169	0,169
Norte	-0,186	-0,213	-0,111	-0,065	-0,180	-0,215	-0,213
Nordeste	-0,348	-0,344	-0,293	-0,293	-0,354	-0,357	-0,346
Sul	-0,042	-0,046	-0,054	-0,015	-0,021*	-0,032	-0,046
Centro	-0,083	-0,089	-0,051	0,032	-0,071	-0,080	-0,088
Formal	0,119	0,174	0,117	0,341	0,136	0,190	0,175
Sind	0,175	0,189	0,263	0,250	0,171	0,185	0,188
Grupo 1**	—	—	—	—	—	—	4,314
Grupo 2	—	—	—	—	—	—	4,323
Grupo 3	—	—	—	—	—	—	4,556
R² ajustado	0,430	0,450	—	—	—	—	0,991
Amostra	15.704	11.1074	40.070	40.070	15.704	111.074	111.074

*: variáveis estatisticamente não significantes ao nível de 5% de significância. A exclusão de tais variáveis pouco afeta os resultados apresentados. As demais variáveis são todas significantes ao nível de 5%. **: são os efeitos individuais, que foram classificados da seguinte maneira: Grupo 1: pessoas entre 24 e 34 anos de idade; Grupo 2: com idade entre 35 e 45 anos; e Grupo 3: pessoas entre 46 e 56 anos de idade.

A análise para o retorno em escolaridade não pode ser feita de forma direta a partir da simples inspeção do coeficiente da variável S devido à existência de produtos cruzados em todos os modelos. De modo então facilitar a análise comparativa do retorno em escolaridade que deriva das quatro metodologias utilizadas, foi montada a tabela 4 que leva em consideração o cálculo dos efeitos cruzados tal como aparece em Garen (1984). Esta tabela mostra os retornos em escolaridade a partir dos sete modelos estimados. Como é interessante observar a variação do

retorno com relação ao nível de escolaridade, a tabela 4 foi dividida em 4 faixas: a) 4 anos de estudo; b) 8 anos de estudo; c) 12 anos de estudo; e d) 16 anos de estudo. De modo ainda a facilitar a elaboração da tabela 4, foram levados em consideração apenas os homens brancos. Por exemplo, no modelo de Garen (modelo 5) estimado com dados em cross-section, para a média dos indivíduos com quatro anos de estudo, a experiência é de 20,63 anos e $\hat{\eta}$ igual a $-5,39$. Tais valores geraram uma taxa de retorno da escolaridade de 0,1189. Já para a média dos indivíduos com oito anos de estudo, a experiência é de 18,99 anos e $\hat{\eta}$ de $-1,49$, o que gera uma taxa de retorno em escolaridade de 0,1555.¹¹

Usando o procedimento de Heckman (1979), com dados empilhados, temos que o retorno em educação, para 4 anos de estudo, é de 0,1449. Usando esta mesma faixa de educação, mas fazendo uso de uma estimativa por pseudo painel (modelo 7), com dados empilhados, o retorno é de 0,1622. Em comparação, o resultado obtido por OLS (modelo 2) para o mesmo conjunto de dados resultou num retorno em escolaridade de 0,1686. Outro detalhe que chama a atenção na tabela 4, é a mudança no retorno em escolaridade à medida que o indivíduo aumenta seu nível educacional. Por exemplo, nas estimativas que adotaram o procedimento de Garen (modelos 5 e 6) o retorno da escolaridade para indivíduos com 16 anos de estudo é quase sempre o dobro do retorno obtido para um nível de escolaridade de 4 anos.

Tabela 4
Retornos da Escolaridade para Homens Brancos

Modelo/Amostra ¹	Retorno em escolaridade para:			
	S = 4	S = 8	S = 12	S = 16
(1) OLS/C	0,1664	0,1680	0,1705	0,1715
(2) OLS/P	0,1686	0,1699	0,1727	0,1733
(3) HECKMAN/C	0,1279	0,1430	0,1470	0,1484
(4) HECKMAN/P	0,1449	0,1583	0,1619	0,1628
(5) GAREN/C	0,1189	0,1555	0,1927	0,2286
(6) GAREN/P	0,1056	0,1419	0,1802	0,2160
(7) EF. FIXO/PP	0,1622	0,1647	0,1703	0,1716
Média	0,1529	0,1644	0,1761	0,1840
Mediana	0,1617	0,1654	0,1727	0,1757
D-P	0,0361	0,0274	0,0256	0,0276

* C = Cross-Section, P = Pooling, PP = Pseudo painel.

¹¹Por motivos de economia de espaço, os valores médios da experiência e de η para cada estimativa não foram divulgados. Entretanto, elas podem ser obtidas diretamente com os autores.

5. Conclusões

Este trabalho teve como objeto estudar a questão do retorno em escolaridade no Brasil. Alguns problemas podem surgir no sentido de obstruir a obtenção de uma estimativa fidedigna para este parâmetro que deve se gerado por meio de uma estimativa correta para a equação de salários. Neste trabalho foram analisados três destes possíveis problemas. O primeiro deles tratado a partir do procedimento de Heckman (1979) visou investigar a existência de viés de seleção causado pela estratégia de “job search” por parte do agente. O segundo problema analisado através de uma metodologia desenvolvida por Garen (1984) se refere a endogeneidade na escolha dos anos de escolaridade. Por fim, o viés de variável omitida, decorrente da ausência de um meio adequado que possa ser usado para medir a habilidade do indivíduo, foi tratado com base na metodologia de pseudo painel (Deaton, 1985).

Para verificar como se comportavam os parâmetros frente a um possível ganho de informação decorrente do aumento da amostra, as estimativas foram feitas tanto para dados *cross-section* do ano 1996, como para dados empilhados (*pooling*) de 1992 a 1999. Uma intuição interessante desses dois conjuntos de dados distintos está em associar a amostra em *cross-section* ao curto prazo, enquanto a amostra em forma de “*pooling*” poderia estar relacionada ao que se observa no longo prazo. Os resultados mostraram que quando se compara o modelo OLS com o de Heckman na tabela 3, observa-se que o efeito do viés de seleção de amostra no curto prazo sobre o retorno em escolaridade, o mesmo não ocorrendo no longo prazo. A constatação desse fato tem um certo apelo, na medida que a imposição de um salário de reserva deve agir com mais força no curto prazo, tendendo a desaparecer num período mais longo do tempo exatamente pela escassez de oferta de trabalho.

No que se refere à análise entre o modelo OLS e o de Garen, observa-se aqui uma regularidade interessante. Primeiro, ela assinala claramente a existência de endogeneidade devido à escolha dos anos de escolaridade. Segundo, tanto para dados em *cross-section* como para dados empilhados, verifica-se uma ultrapassagem do retorno em escolaridade a partir da faixa de 12 anos de estudo em relação ao retorno gerado por OLS, o que indica uma mudança no viés gerado por endogeneidade na decisão por grau de escolaridade. Isso significa que para níveis baixos de escolaridade, abaixo do *breaking point*, o retorno é bem reduzido, o que não ocorre para níveis acima desse ponto. Assim, se tomarmos por base a equação de determinação individual dos anos de estudo (equação de educação) isso significa que por sua própria iniciativa, o agente tem uma baixa propensão em investir em escolaridade se ele se encontra num nível corrente de escolaridade abaixo do

breaking point, pelo fato da sua expectativa de retorno ser muito baixa. Caso seu nível de escolaridade esteja acima deste, o seu incentivo de investir em educação é bastante elevado devido ao alto retorno. Como as estimativas geradas por OLS não levam em consideração o fato da decisão por escolaridade ser endógena, não existe basicamente mudança no retorno para as quatro faixas de estudos presentes na tabela 4.

Uma consequência direta do que acabou de ser posto, diz respeito ao fato de que no caso da participação do governo no setor de educação, isso deve ser feito principalmente para níveis mais elementares de ensino. Uma pessoa de baixa renda e com pouca escolaridade não estaria propensa a investir em educação, pois teria que suportar um custo muito alto gerado pelo baixo retorno até que o investimento se tornasse atrativo. O mesmo raciocínio vale para famílias de baixa renda, que pelo mesmo motivo não estariam dispostas a investir na escolaridade dos filhos.

Após essas considerações, quatro outras conclusões podem ainda ser derivadas deste estudo: 1) foram encontradas evidências que suportam a hipótese de vantagens comparativas proposta por Willis e Rosen (1979); 2) tal como em Garen (1984) os retornos em escolaridade parecem variar com o nível de escolaridade do indivíduo, fato que também aparece em Katz e Autor (1999); 3) a similaridade entre os resultados obtidos por OLS e por pseudo painel para o retorno em escolaridade tal como se verifica na tabela 4 parece indicar que o efeito de variável omitida sobre o retorno em escolaridade é nulo ou pouco perceptível, tal como aparece em outros estudos; e 4) por fim, os resultados indicam ainda a existência de endogeneidade na escolha da escolaridade, assim como a presença de viés de seleção no curto prazo advinda da estratégia de procura por emprego.

Referências

- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70:9–49.
- Becker, G. S. (1975). *Human Capital: A Theoretical and Empirical Analysis*. Columbia University Press, New York.
- Bratsberg, B. & Terrell, D. (2002). School quality and returns to education of U.S. immigrants. *Economic Inquiry*, 40(2):177–98.
- Cameron, S. V. & Heckman, J. (2001). The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy*, 109(3):455–99.

- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.
- Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30:109–126.
- Dickens, W. & Lang, K. (1985). A test for dual labor market theory. *American Economic Review*, 75:792–805.
- Dickens, W. & Lang, K. (1992). Labor market segmentation theory: Reconsidering the evidence. NBER Working Paper 4087.
- Garen, J. (1984). The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica*, 52(5):1199–1218.
- Griliches, Z. (1977). Estimating the returns to schooling: Some econometrics problems. *Econometrica*, 45(1):1–22.
- Gronau, R. (1974). Wage comparisons: A selectivity bias. *Journal of Political Economy*, 82(6):1119–1155.
- Hausman, J. A. & Taylor, W. (1981). Panel data and unobservable individual effects. *Econometrica*, 49(6):1377–1398.
- Heckman, J. (1974). Shadow prices, market wages and labor supply. *Econometrica*, 42(4):679–694.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.
- Heckman, J., Tobias, J. L., & Vytlačil, E. (2000). Simple estimators for treatment parameters in a latent variable framework with an application to estimating the returns to schooling. NBER Working Paper, 7950.
- Kassouf, A. L. (1994). Wage rate estimation using heckman procedure. *Revista de Econometria*, XIV(1):89–107.
- Katz, L. & Autor, D. (1999). Changes in the wages structure and earnings inequality. In Ashenfelter, O. & Card, D., editor, *Handbook of Labor Economics*. North Holland. Volume 3A.
- Kenny, L., Lee, L. F., Maddala, G. S., & Trost, R. P. (1979). Returns to college education: An investigation of self-selection bias based on the project talent data. *International Economic Review*, 20:775–89.

- Loureiro, P. R. A. & Galvão, F. G. (2001). Discriminação no mercado de trabalho: Uma análise dos setores rural e urbano no Brasil. *Economia Aplicada*, 5(3):519–545.
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, 66(4):281–302.
- Mincer, J. (1974). Schooling, experience and earnings. National Bureau of Economic Research.
- Silva, N. D. V. & Kassouf, A. L. (2000). Mercados de trabalho formal e informal: Uma análise da discriminação e da segmentação. *Nova Economia*, 10(1):41–77.
- Soares, R. R. & Gonzaga, G. (1999). Determinação de salários no Brasil: Dualidade ou não-linearidade no retorno à educação. *Revista de Econometria*, 19(2):377–404.
- STATA (2003). *Stata User's Guide*. Stata Press, Texas. Realease 8, College Station.
- Tobias, J. L. (2003). Are returns to schooling concentrated among the most able? a semiparametric analysis of the ability-earnings relationships. *Oxford Bulletin of Economics and Statistics*, 61(1):1–29.
- Topel, R. H. (1999). Labor markets and economic growth. In Ashenfelter, O. & Card, D., editor, *Handbook of Labor Economics*. North Holland. Volume 3C.
- Ueda, E. M. & Hoffmann, R. (2002). Estimando o retorno em educação no Brasil. *Economia Aplicada*, 6(2):209–238.
- Willis, R. J. (1986). Determinants: A survey and reinterpretation of human capital earnings functions. In *Handbook of Labor Economics*. Elsevier Science Publishers, BV. V. I.
- Willis, R. J. & Rosen, S. (1979). Education and self selection. *Journal of Political Economy*, 87(5):S7–S36.