

RETORNOS PARA EDUCAÇÃO NO BRASIL: EVIDÊNCIAS EMPÍRICAS ADICIONAIS*

Marcelo Resende[§]
Ricardo Wyllic[□]

RESUMO

O artigo investiga econometricamente os retornos para educação no Brasil, tomando como referência dados da Pesquisa sobre Padrão de Vida (PPV-IBGE) antes não explorados nesse contexto. Equações de rendimentos para homens e mulheres são consideradas mediante o procedimento para correção do efeito de seletividade amostral concebido por Heckman, incluindo uma medida de qualidade para educação. Os resultados obtidos apontam para retornos semelhantes à literatura anterior.

Palavras-chave: retornos para educação, qualidade, seletividade amostral.

ABSTRACT

The paper econometrically investigates returns for education in Brazil, taking as reference data from the Pesquisa sobre Padrão de Vida (PPV-IBGE) previously not explored in that context. Earnings equations are considered for men and women in terms of the sample selection correction procedure conceived by Heckman, including a measure for quality in education and a more direct measurement of experience. The obtained results indicate similar returns as compared to the previous literature.

Key words: returns to education, quality, sample selection.

JEL classification: J20, J24.

* Os autores agradecem os comentários de um parecerista anônimo, mas os eventuais erros remanescentes são de nossa exclusiva responsabilidade.

§ Instituto de Economia-Universidade Federal do Rio de Janeiro. Av. Pasteur 250, Urca, 22290-240, Rio de Janeiro-RJ. Email: mresende@ie.ufrj.br.

□ Fundação Centro de Informações e Dados do Rio de Janeiro (CIDE). Rua Pinheiro Machado s/n - Palácio Guanabara – Prédio Anexo IV. 22231-901, Rio de Janeiro - R.J. E-mail: rwyllie@cide.rj.gov.br.

Recebido em janeiro de 2005. Aceito em janeiro de 2006.

1 INTRODUÇÃO

Estudos econométricos acerca dos retornos para educação têm sido recorrentes em diferentes países (ver Psacharopoulos, 1981 e Ashenfelter *et al.*, 1999). Com efeito, as fontes de crescimento econômico têm sido crescentemente associadas a investimentos em capital humano (ver, por exemplo, Denison, 1974) e o fenômeno da decolagem dos chamados tigres asiáticos esteve amplamente relacionado a investimentos expressivos em educação.

Em economias como a brasileira, que é caracterizada por uma das mais elevadas desigualdades na distribuição de renda, a investigação da magnitude dos retornos de educação reveste-se de especial interesse para um possível enfrentamento do problema. Neste sentido, pode-se mais recentemente destacar os estudos de Lam e Levison (1990), Leal e Werlang (1991), Kassouf (1994, 1998) e Sachsida *et al.* (2004). Uma influente linha de investigação segue o trabalho seminal de Mincer (1974), que concebeu uma equação para rendimentos que seria dependente de fatores explicativos associados à escolaridade e à experiência (excetuando outras variáveis de controle). As contribuições de Kassouf e Sachsida *et al.* inspiram-se em tal abordagem, mas procuram levar em consideração um instrumental econométrico mais atualizado para abarcar a questão da seletividade amostral.

Um aspecto pouco investigado na literatura sobre retornos de educação refere-se à qualidade da mesma. De fato, Behrman e Birdsall (1983) e Behrman *et al.* (1996) salientam que a ausência de controles para esse fator pode levar a estimativas distorcidas para países em desenvolvimento, marcados por amplas heterogeneidades, e chegam a citar especificamente o caso brasileiro.

O presente trabalho investiga a magnitude dos retornos para educação (para os subgrupos de homens e mulheres) e pretende contribuir para a literatura sob três aspectos, a saber:

- a) Avaliar a robustez de evidência empírica anterior em termos de uma base de dados pouco explorada, qual seja, a Pesquisa sobre Padrão de Vida (PPV-IBGE);
- b) Aprimorar a medida de experiência dos trabalhadores diante da disponibilidade de informação sobre a idade de ingresso no mercado de trabalho;
- c) Viabilizar a construção de uma variável de qualidade de educação a partir da percepção dos indivíduos da amostra.

O artigo encontra-se organizado da seguinte forma. A segunda seção apresenta uma breve digressão acerca das diferentes abordagens para a quantificação dos retornos para educação. A terceira seção discute os dados utilizados e apresenta os resultados da estimação econométrica. A quarta seção apresenta os comentários finais e sugestões para pesquisas futuras.

2 RETORNOS PARA EDUCAÇÃO: UMA DIGRESSÃO

Nesta seção apresentamos de forma sumária os principais elementos envolvidos na quantificação de retornos para educação. De forma esquemática, podem ser destacadas **quatro** fases principais da evolução desses estudos (ver Willis, 1986 e Berndt, 1991):¹

- a) **Estudos não econométricos** (ver, por exemplo, Mincer, 1958, 1962). Esta linha de pesquisa motivou estudos nos quais são identificados custos de educação e rendimentos do trabalho com vistas a calcular a taxa interna de retorno. Tal abordagem peca pela dificuldade em efetuar o

1 Vale ressaltar que além dessa trajetória de desenvolvimentos metodológicos existiram esforços de melhoria nos dados utilizados. Com efeito, uma crítica aos estudos de retornos para educação está associada à possível dificuldade de se utilizar de habilidades não observadas como variáveis de controle, e é nessa linha que surgiram estudos com gêmeos. Adicionalmente, a própria qualidade do esforço educativo pode ser não homogênea, o que motiva desenvolvimento de bases de dados mais abrangentes.

controle para outras variáveis determinantes dos rendimentos, conforme seria mais prontamente atingível em uma análise econométrica.

- b) **Estudos econométricos simplificados baseados em estimação de mínimos quadrados ordinários.** Esta linha, essencialmente motivada pela teoria do capital humano, tem em Mincer (1974) uma contribuição-chave. O mesmo propõe uma equação da forma:

$$\ln y_i = f(s_i, x_i, z_i) + u_i \quad i = 1, \dots, n \quad (1)$$

em que $\ln y_i$ denota o logaritmo natural dos rendimentos do indivíduo i , s_i é uma medida de escolaridade, x_i denota uma variável representativa da experiência e u_i é um erro estocástico. Esse autor e outros trabalhos sucedâneos tipicamente consideraram uma especificação linear para $f(\cdot)$ e também variantes que por vezes incluem termos quadráticos nas variáveis de escolaridade e experiência refletindo a possibilidade de retornos decrescentes naquelas variáveis. Alternativamente, alguns estudos recorrem ao emprego de *dummies* para a escolaridade como estratégia para capturar uma possível não-linearidade dos retornos. A inclusão de variável de experiência pretende enfatizar a importância do treinamento no trabalho e da aprendizagem com a experiência (*learning by doing*). Cabe ressaltar que uma expressão com (1) deve ser pensada como uma aproximação para o problema intertemporal no qual o indivíduo escolhe trajetórias de carreira e formação de capital humano de forma a maximizar a riqueza sujeita à função-utilidade e restrição orçamentária intertemporal. Com efeito, as tentativas de obter soluções fechadas para esse tipo de problema não têm obtido sucesso (ver Willis, 1986). Do ponto de vista econométrico, essa literatura inicial ignora o problema de seletividade amostral. De fato, o estimador de mínimos quadrados ordinários quando aplicado a uma equação de retorno que considere somente indivíduos empregados (com renda positiva) é inconsistente.²

- c) **Estudos econométricos que tratam da endogeneidade e da omissão de explicativas**, ambas relacionadas à escolaridade.

A preocupação com a presença de endogeneidade na variável “anos de estudo”, e seus reflexos sobre a formulação dos modelos, tem origem motivada no trabalho de Backer (1975), que destaca a perspectiva do retorno para educação como fator relevante na decisão de investir e, conseqüentemente, na determinação do próprio nível de escolaridade. Em termos mais práticos, cabe mencionar as metodologias desenvolvidas por Willis (1979) e Garen (1984), que controlam, por meio de agrupamentos de dados e da inclusão de uma equação específica, respectivamente, os efeitos de fatores relacionados com os rendimentos do trabalho.

Um outro problema que tem merecido muita atenção diz respeito à não inclusão de variáveis explicativas relevantes, não influenciadas pelos retornos, mas que, em tese, estariam correlacionadas com a variável “anos de estudo”. Neste particular, o “viés de habilidade” – inteligência – e a qualidade da educação têm estimulado os estudiosos na proposição de soluções que contornem a tendenciosidade e a inconsistência dos estimadores para o retorno. A preocupação com o controle para fatores específicos, associados ao desempenho individual, fica evidenciada pela utilização de amostras com gêmeos univitelinos, citada em Hoffmann e Ueda (2002), e diretamente no trabalho desenvolvido por Lam e Schoeni (1993), que contemplam a estrutura familiar. Outra fonte de viés está na qualidade da educação, também objeto de muitos esforços, como em Behrman e Birdsall (1983) e Behrman *et al.* (1996), Birdsall e Kaplan (1996).

2 A abordagem de Mincer tem sido objeto de críticas. Heckman *et al.* (2003) discutem as condições segundo as quais o coeficiente da variável de escolaridade na equação de rendimentos possui interpretação não ambígua. Os autores sugerem o uso preferencial de procedimentos não paramétricos e destacam que fatores importantes para explicar o retorno educacional são usualmente ignorados. Merecem menção os custos diretos e indiretos para adquirir educação, a estrutura de tributação, a duração da vida profissional e a incerteza sobre retornos futuros na ocasião da decisão educacional.

d) **Estudos econométricos lidando com o problema da seletividade amostral.** Os elementos metodológicos dessa abordagem aprimorada foram apresentados por Heckman (1974, 1979). O procedimento em dois estágios pode ser sumariado da seguinte forma: i) considere a estimação de um modelo Probit para a participação no mercado de trabalho (levando em conta, portanto, uma amostra com trabalhadores empregados e desempregados), a partir da qual se pode obter a razão inversa de Mills para cada ponto da amostra; ii) considere a estimação por mínimos quadrados de uma equação de retorno (por exemplo, à Mincer) para a amostra de trabalhadores empregados que inclua como regressor adicional a razão inversa de Mills. Pode-se mostrar que os erros da equação de retorno seriam heteroscedásticos. Neste sentido, é desejável utilizar erros padrões robustos à heteroscedasticidade na linha de White.³

Tendo fornecido um breve panorama das abordagens para obtenção de retornos para educação, cumpre comentarmos brevemente as contribuições mais recentes no caso brasileiro.

Lam e Levison (1991) estudam a prevalência de diferenciais de renda nos E.U.A. e no Brasil e sua relação com idade, experiência e escolaridade. Para o caso brasileiro, os autores fazem uso de dados da Pesquisa Nacional por Amostra de domicílios (PNAD-IBGE) de 1985. Foram consideradas regressões bastante simplificadas do logaritmo natural do rendimento em função da escolaridade, que foram estimadas para diferentes faixas de experiência e idade. Os retornos assim obtidos foram elevados (tipicamente acima de 17%). Todavia, como os outros fatores explicativos relevantes não são controlados diretamente no modelo de regressão considerado, a análise em questão serve mais para detalhar o papel de variáveis de experiência e idade do que propriamente fornecer um valor para a taxa de retorno para educação.

Leal e Werlang (1991) consideraram um modelo na linha da equação de rendimentos de Mincer tomando como referência dados da Pesquisa Nacional por Amostra de domicílios (PNAD-IBGE) para o período 1976-89. A equação em questão foi estimada para cada um dos anos do período retromencionado. As taxas de retorno obtidas foram elevadas, chegando a ultrapassar 15% em alguns casos. Uma limitação importante destes dois últimos estudos refere-se ao não tratamento da questão da seletividade amostral.

Kassouf (1994, 1998) avança na literatura brasileira ao considerar a questão da seletividade amostral. Os dois trabalhos fazem uso de dados da Pesquisa Nacional de Saúde e Nutrição. (IBGE, 1989). São consideradas equações de rendimentos para homens e mulheres. O primeiro trabalho procura evidenciar os vieses decorrentes da adoção de procedimentos econométricos tradicionais que não corrigem para a seletividade amostral. Os resultados obtidos apontam para retornos moderados (7,2% e 8,5% para homens e mulheres, respectivamente). No segundo artigo centraliza sua atenção na discriminação entre sexos. Todavia, este procedimento apresenta uma interessante extensão do procedimento para correção da seletividade amostral que pode ser relevante em estimações de retornos para educação em geral. No primeiro estágio do procedimento de Heckman, em vez de utilizar um modelo Probit para a equação de participação no mercado de trabalho, a autora lança mão de um modelo Logit multinomial com uma terceira opção de informalidade na variável dependente (adicionalmente ao desemprego e emprego formal). A segunda etapa do procedimento segue normalmente com a inclusão da razão inversa de Mills como regressor adicional na equação de rendimentos.

O estudo de Sachsida *et al.* (2004) utiliza a Pesquisa Anual por Amostra de Domicílios-PNAD durante o período 1992-99. Os autores desenvolvem estimativas em *cross-section* para o ano de 1996 e com uma estrutura de pseudopainel para todo o período. Similarmente a Kassouf, os

3 Kassouf (1994) aplica o procedimento de dois estágios de Heckman para o caso brasileiro e fornece uma apresentação sumária útil do mesmo.

autores também consideraram o problema da seletividade amostral e procuram tornar endógena a escolaridade na análise. Em seu modelo preferencial (procedimento de Heckman com pseudopainel) os autores obtêm retornos para educação entre 16% e 17,5% para os homens.

O presente estudo segue uma linha semelhante à de Kassouf (1994) ao abordar o problema da seletividade amostral no contexto da estimação de retornos para educação a partir de dados da Pesquisa sobre Padrão de Vida (PPV-IBGE). Além da consideração de base de dados pouco explorada nesse contexto, o artigo pretende contribuir para a literatura sob dois aspectos, a saber:

- a) **Consideração de uma medida mais precisa da variável de experiência.** Com efeito, virtualmente a totalidade da literatura supramencionada adota a tradicional medida que desconta da idade a escolaridade e o ano modal de ingresso na escola (6 anos), ignorando a possibilidade de contagem simultânea de anos de estudo e experiência de trabalho. Na PPV, contudo, tem-se informação sobre a idade de ingresso no mercado de trabalho que, subtraída da idade do indivíduo, pode gerar uma medida mais precisa de experiência. A exemplo das anteriores, esta medida também não estará livre de vies por conta de períodos de inatividade.
- b) **Qualidade da educação como variável de controle.** A importância deste tipo de controle já fora destacada no trabalho de Behrman e Birdsall (1983). Com efeito, tal trabalho indica que a desconsideração do fator qualidade tenderia a levar a uma superestimação do retorno para educação. Os autores implementaram uma ilustração empírica para o Brasil tomando como base o Censo de 1970. Adotou-se como *proxy* para qualidade o número médio de anos de estudo de professores na área na qual o indivíduo realizou sua educação (medida por Estado). Assim procedendo, os resultados indicam um amplo vies para cima ao se omitir a variável para qualidade, obtendo-se um retorno de 20,5% em contraste com o retorno de 4,7% para os homens na especificação preferencial que inclui a qualidade.

Cumpra observar que a evidência deve ser encarada com cautela, já que o estudo baseou-se em uma amostra restrita para homens entre 15 e 35 anos, de modo a garantir maior homogeneidade da *proxy* para qualidade. Diferentes críticas podem ser levantadas a tal *proxy* de qualidade, e que são arroladas pelos próprios autores:

- i) A agregação por estado pode gerar importantes erros de medida;
- ii) Supõe-se que a qualidade da educação recebida pelos professores é uniforme;
- iii) A qualidade dos professores é um dos aspectos importantes, mas infra-estrutura e qualidade do material didático também são essenciais;
- iv) A *proxy* para qualidade considera o nível educacional médio dos professores no ano da base de dados, mas a educação recebida pelos indivíduos se deu em períodos anteriores;
- v) Migrações múltiplas podem tornar difícil a identificação do local onde a educação foi recebida;
- vi) A educação formal é um dos componentes de um ambiente de aprendizado no qual o grau de instrução das pessoas próximas e o acesso a livros e meios culturais em geral desempenham um papel importante.

Behrman *et al.* (1996) consideram uma medida semelhante, a escolaridade média dos professores da localidade na qual o indivíduo cursou as quatro primeiras séries do ciclo fundamental de ensino (antigo primário). No âmbito da disponibilidade de dados da PPV a construção dessa medida é inviável. Alternativamente, na PPV estão disponíveis dados sobre a qualidade da educação recebida pelos indivíduos a partir da opinião dos mesmos. Aqui, esta irá servir como variável para a qualidade da educação, buscando evitar que o retorno apresente vies em razão de uma possível

correlação entre o número de anos de estudo e a qualidade da educação. Porém, não restam dúvidas quanto à possibilidade de que a opinião dos indivíduos sobre a sua própria educação possa estar, pelo menos em parte, refletindo a satisfação dos mesmos com seus rendimentos do trabalho. Contudo, tal dificuldade mostra-se recorrente na literatura. Com efeito, conforme Becker (1975), a escolaridade dos professores primários proposta em Behrman e Birdsall (1983) deverá depender do retorno esperado e, portanto, dos rendimentos projetados. Estes, por sua vez, irão variar de acordo com os rendimentos da região em que se encontram radicados. Uma alternativa não implementada no presente estudo poderia ser a inclusão de variável de qualidade em termos de equação considerando endogeneidade nos anos de estudo (ver Garen, 1984). Segue-se a análise empírica.⁴

3 ANÁLISE EMPÍRICA

3.1 Base de dados

O presente trabalho toma como base a PPV. Durante os anos de 1996 e 1997 foram realizadas entrevistas no âmbito das regiões Sudeste e Nordeste. A pesquisa abrange 10 áreas que incluem: áreas metropolitanas de Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, outras áreas urbanas da região Sudeste e áreas rurais das regiões Sudeste e Nordeste.

O desenho amostral é semelhante aos adotados em outras pesquisas do IBGE, como na PNAD, por exemplo. São dois os estágios de seleção, definindo-se no primeiro os setores censitários como unidades amostrais, organizados em estratos, e no segundo os domicílios. Foram selecionados 544 setores, visitados 4.940 domicílios e coletadas informações sobre 19.409 indivíduos. Muito embora a amostra não seja do tipo autoponderada, os procedimentos aqui adotados para a utilização dos microdados da pesquisa estão de acordo com os empregados na literatura em geral. Trata-se de uma base ainda pouco explorada, merecendo menção os trabalhos de Barros *et al.* (1999) e Ferreira *et al.* (2000). O primeiro considera a estrutura salarial (especialmente no que tange a diferenciais de salário), enfatizando o papel de características do trabalhador e do posto de trabalho e possui um caráter descritivo. O segundo trabalho considera essa base de dados e ainda a PNAD para traçar um perfil mais pormenorizado da pobreza e enfatiza a importância de aspectos espaciais e do nível educacional na configuração daquela. A despeito desses esforços, a base da PPV permite explorações adicionais como, por exemplo, o estudo econométrico dos retornos para educação que se empreende a seguir.

As Tabelas 1 e 2 apresentam as estatísticas descritivas e as descrições das variáveis utilizadas para os subgrupos de homens e mulheres. Na maioria dos casos as descrições são auto-explicativas, cabendo, contudo, algumas observações. Foram utilizadas variáveis *dummy* de intercepto para raça, região, tipo da área (metropolitana ou rural) e posição no domicílio. No caso da raça, excluiu-se a amarela. No caso da região foi excluída a variável *dummy* referente ao Sudeste, e finalmente as categorias rural e outra posição no domicílio que não as explicitadas (essencialmente pessoas sem relação de parentesco e sem relação mais direta) para as variáveis de área e de posição no domicílio, respectivamente.

4 Card e Krueger (1992) já haviam enfatizado a importância de incluir variáveis que reflitam, de forma mais direta, a qualidade da educação, tendo utilizado, em particular, a relação professores/alunos, a duração média do período letivo e medida de salário dos professores.

Tabela 1 – Estatísticas descritivas – mulheres

Variável	Descrição	Média	Desvio Padrão
IDADE	idade em anos	35,73	14,33
EDUC	escolaridade em anos (excluindo repetências)	5,44	4,57
CRIANÇAS	número de crianças entre 0 e 6 anos de idade no domicílio	0,57	0,87
DURB	assume valor 1 se o indivíduo está situado em área urbana e 0 caso contrário	0,72	0,45
DREG	assume valor 1 se o indivíduo está situado na região Nordeste 0 se estiver situado na região Sudeste	0,48	0,50
BRAN	assume valor 1 se o indivíduo for branco e 0 caso contrário	0,47	0,50
NEG	assume valor 1 se o indivíduo for negro e 0 caso contrário	0,07	0,25
PARD	assume valor 1 se o indivíduo for pardo e 0 caso contrário	0,46	0,50
SAL	rendimento do trabalhador (por hora) na atividade principal	2,53	27,77
OUTAPO	outros rendimentos (aposentadoria, previdência privada e pensão) do domicílio	149,38	545,11
EXPE	número de anos de experiência, definida como IDADE-idade de ingresso no mercado de trabalho e como 0 para quem nunca trabalhou	20,96	15,97
CHEFE	assume valor 1 se o indivíduo for chefe de família e 0 caso contrário	0,46	0,50
CONJ	assume valor 1 se o indivíduo for cônjuge e 0 caso contrário	0,25	0,43
FILHA	assume valor 1 se o indivíduo for filho(a) no domicílio e 0 caso contrário	0,24	0,43
OUTRO	assume valor 1 se o indivíduo possuir outra relação de parentesco no domicílio e 0 caso contrário	0,004	0,19
QUAL	assume valor 1 quando a educação recebida foi julgada como muito boa ou boa e 0 caso contrário	0,41	0,49

Tabela 2 – Estatísticas descritivas – homens

Variável	Descrição	Média	Desvio Padrão
IDADE	idade em anos	34,42	15,08
EDUC	escolaridade em anos (excluindo repetências)	5,61	4,73
CRIANÇAS	número de crianças entre 0 e 6 anos de idade no domicílio	0,51	0,83
DURB	assume valor 1 se o indivíduo está situado em área urbana e 0 caso contrário	0,77	0,42
DREG	assume valor 1 se o indivíduo está situado na região Nordeste 0 se estiver situado na região Sudeste	0,50	0,50
BRAN	assume valor 1 se o indivíduo for branco e 0 caso contrário	0,47	0,50
NEG	assume valor 1 se o indivíduo for negro e 0 caso contrário	0,07	0,25
PARD	assume valor 1 se o indivíduo for pardo e 0 caso contrário	0,46	0,50
SAL	rendimento do trabalhador (por hora) na atividade principal	1,14	18,32
OUTAPO	outros rendimentos (aposentadoria, previdência privada e pensão) do domicílio	188,64	579,34
EXPE	número de anos de experiência, definida como IDADE- idade de ingresso no mercado de trabalho	18,07	16,46
CHEFE	assume valor 1 se o indivíduo for chefe de família e 0 caso contrário	0,29	0,46
CONJ	assume valor 1 se o indivíduo for cônjuge e 0 caso contrário	0,31	0,46
FILHA	assume valor 1 se o indivíduo for filho(a) e 0 caso contrário domicílio e 0 caso contrário	0,32	0,47
OUTRO	assume valor 1 se o indivíduo possuir outra relação de parentesco no	0,06	0,24
QUAL	assume valor 1 quando a educação recebida foi julgada como muito boa ou boa e 0 caso contrário	0,42	0,49

3.2 Resultados

Nesta seção são apresentados os resultados das estimações obtidas pelo procedimento de dois estágios de Heckman. Os resultados iniciais para as equações de participação no mercado de traba-

lho para homens e mulheres são reportados na Tabela 3.⁵ Para tais equações não existe uma especificação consensual na literatura. Aqui seguimos de perto a literatura, com exceção de uma variável relativa a outros rendimentos e da variável “crianças”. No primeiro caso adota-se uma definição distinta, ao passo que no segundo caso evita-se um tratamento assimétrico associado à exclusão da variável da equação de homens.⁶

Tabela 3 – Equação de participação no mercado de trabalho

Variáveis	Mulheres	Homens
Constante	-2,075 (0,000)	-3,194 (0,000)
EDUC	0,140 (0,000)	0,161 (0,000)
EDUC*2	-0,004 (0,000)	-0,003 (0,002)
IDADE	0,026 (0,000)	0,044 (0,000)
IDADE*2	-0,001 (0,000)	-0,001 (0,000)
CRINANÇAS	0,069 (0,001)	0,034 (0,187)
CHEFE	0,746 (0,000)	0,442 (0,008)
CONJ	0,309 (0,076)	0,434 (0,009)
FILHA	0,269 (0,116)	0,133 (0,414)
OUT	0,307 (0,104)	0,191 (0,282))
DURB	-0,116 (0,003)	-0,008 (0,874)
DREG	-0,279 (0,000)	-0,121 (0,003)
BRAN	0,460 (0,150)	0,672 (0,081)
NEG	0,828 (0,011)	0,764 (0,052)
PARD	0,695 (0,030)	0,808 (0,037)
OUTAPO	-0,300 (0,000)	-0,000 (0,000)
Previsão correta	68,15 %	76,84 %
Número de observações	6751	5998

A inspeção da tabela mostra que obtivemos coeficientes predominantemente significativos e com sinais adequados. Exceções surgem com as *dummies* para posição no domicílio e de raça. Adicionalmente, a variável CRIANÇAS não teve um coeficiente significativo para o caso dos homens, justificando o tratamento diferenciado sugerido por Kassouf.

5 A estimação de modelos separados para homens e mulheres decorre do consenso de que as diferenças nesses casos não podem ser acomodadas pela simples introdução de variáveis *dummies* de intercepto (ver Cairns, 1986).

6 A referida questão é importante sobretudo em um contexto no qual a participação da mulher no mercado de trabalho torna-se progressivamente mais importante. Ver Sedlacek e Santos (1991).

A Tabela 4 apresenta os resultados das equações de rendimento para mulheres e homens. Nesse segundo estágio tomamos como referência especificações mais difundidas de Mincer. Especificamente:

$$\ln SAL_i = \beta_0 + \beta_1 EDUC_i + \beta_2 EXP_i + \beta_3 EXP_i^2 + \beta_4 \lambda_i + \gamma.Z_i + u_i \quad (2)$$

em que λ_i refere-se à razão inversa de Mills usada para considerar a questão da seletividade amostral, Z_i denota um vetor de outras variáveis de controle relevantes e γ o vetor de parâmetros correspondentes. Essa equação indica genericamente o papel da variável experiência. Neste sentido, a formulação de Kassouf, que inclui apenas as variáveis relativas à idade, é potencialmente limitada. É desejável uma exploração adicional desse aspecto em termos de uma variável de experiência construída conforme a discussão da seção anterior.

Tabela 4 – Equação de retorno salarial (15-70 anos) com a razão de Mills

Variáveis Explicativas	Mulheres		Homens	
	Com qualidade	Sem qualidade	Com qualidade	Sem Qualidade
Constante	0,430 (0,277)	0,518 (0,192)	-1,406 (0,000)	-1,565 (0,000)
EDUC	0,126 (0,000)	0,135 (0,000)	0,159 (0,000)	0,174 (0,000)
DURB	0,573 (0,000)	0,593 (0,000)	0,338 (0,000)	0,353 (0,000)
DREG	-0,355 (0,000)	-0,371 (0,000)	-0,379 (0,000)	-0,389 (0,000)
BRAN	-1,016 (0,002)	-1,089 (0,001)	-0,114 (0,492)	-0,065 (0,689)
NEG	-1,209 (0,000)	-1,298 (0,000)	-0,379 (0,034)	-0,357 (0,044)
PARD	-1,276 (0,000)	-1,356 (0,000)	-0,269 (0,121)	-0,231 (0,177)
QUAL	0,264 (0,000)	-	0,256 (0,000)	-
EXPE	0,036 (0,000)	0,379 (0,000)	0,041 (0,000)	0,043 (0,000)
EXPE*2	-0,001 (0,000)	-0,0005 (0,000)	-0,001 (0,000)	-0,0007 (0,000)
LAMBDA	-0,329 (0,014)	-0,313 (0,020)	0,343 (0,013)	0,437 (0,002)
R ² ajustado	0,487	0,479	0,477	0,465
Número de observações	2660		1621	

A significância do coeficiente relativo à λ nas diferentes equações corrobora a importância do procedimento de correção para seletividade amostral. Os resultados obtidos, quer em termos da significância dos diferentes coeficientes, quer dos sinais obtidos, são em geral bastante bons. Vale destacar os resultados obtidos a partir de nossa medida mais direta de experiência que, além de

significativas, revelaram-se estar em um patamar inferior aos valores obtidos por outros autores.⁷ Adicionalmente, convém ressaltar a importância de se usar a qualidade da educação como variável de controle. Os sinais positivos dos coeficientes, tanto para o caso de homens quanto de mulheres, estão de acordo com as expectativas, sendo ambos significativos.

As estimativas dos retornos em educação foram de 12,6% e 15,9% para mulheres e homens, respectivamente. Fica claro que as estimativas encontradas estão próximas daquelas obtidas nos estudos anteriormente mencionados, com exceção de Kassouf (1994), que além de obter valores bem mais baixos encontra evidência de retornos mais baixos para homens.

Embora já fique clara a importância da correção para seletividade amostral, a Tabela 5 apresenta, para fins comparativos, estimativas sem a correção. Adicionalmente, as duas tabelas procuram verificar a sensibilidade dos resultados relativamente à inclusão da variável de qualidade. Nossas estimativas preferenciais encontram-se, portanto, na Tabela 4.

Tabela 5 – Equação de retorno salarial (15-70 anos) sem a razão de Mills

Variáveis Explicativas	Mulheres		Homens	
	Com qualidade	Sem qualidade	Com qualidade	Sem Qualidade
Constante	-0,076 (0,818)	0,034 (0,917)	-0,609 (0,001)	-0,545 (0,001)
EDUC	0,141 (0,000)	0,150 (0,000)	0,131 (0,0000)	0,139 (0,000)
DURB	0,554 (0,000)	0,575 (0,000)	0,340 (0,000)	0,357 (0,000)
DREG	-0,412 (0,000)	-0,425 (0,000)	-0,346 (0,000)	-0,347 (0,000)
BRAN	-0,953 (0,003)	-1,028 (0,001)	-0,270 (0,086)	-0,264 (0,066)
NEG	-1,059 (0,001)	-1,154 (0,000)	-0,563 (0,001)	-0,596 (0,000)
PARD	-1,160 (0,000)	-1,244 (0,000)	-0,461 (0,004)	-0,477 (0,001)
EXPE	0,043 (0,000)	0,044 (0,000)	0,034 (0,000)	0,033 (0,000)
EXPE*2	-0,0006 (0,000)	-0,0006 (0,000)	-0,0005 (0,000)	-0,0005 (0,000)
QUAL	0,260 (0,000)	-	0,272 (0,000)	-
R ² ajustado	0,486	0,477	0,474	0,461
Número de observações	2660		1621	

Por fim, a Tabela 6 resume os principais resultados advindos das diferentes especificações para a equação de retorno.

7 Os resultados de Kassouf (1998) estão entre 0,005 e 0,10, ao passo que Sachsida *et al.* (2004) encontraram valores entre 0,048 e 0,094.

Tabela 6 – Comparativo dos resultados com e sem a correção de Heckman

Variáveis Explicativas	Mulheres		Homens	
	Com qualidade	Sem qualidade	Com qualidade	Sem qualidade
EDUC (com correção)	0,126 (0,000)	0,135 (0,000)	0,159 (0,000)	0,174 (0,000)
EDUC (sem correção)	0,141 (0,000)	0,150 (0,000)	0,131 (0,000)	0,139 (0,000)
QUAL (com correção)	0,264 (0,000)	-	0,256 (0,000)	-
QUAL (sem correção)	0,260 (0,000)	-	0,272 (0,000)	-
Número de observações	2660		1621	

A não correção para seletividade amostral leva à geração de vieses em direções distintas para homens e mulheres. Efetivamente, o sinal do coeficiente da variável de correção da seletividade amostral (λ) depende de dois parâmetros. Trata-se da correlação entre os erros das equações de participação e de retornos e o sentido de truncamento efetuado sobre os erros da primeira equação. Na aplicação aqui descrita, o truncamento é definido por um limite inferior, o que significa dizer que o sinal do parâmetro será o mesmo da correlação, segundo Greene (1993). Portanto, para as mulheres, o sinal negativo encontrado para a estimativa do parâmetro de λ indica que as variáveis que aparecem incluídas na equação de participação concorrem, em conjunto, para elevar a média condicional dos salários femininos, dando-se o inverso no caso masculino. Do ponto de vista econômico, sinais distintos indicam que as estratégias de busca (*'job search'*) reagem em direções opostas com respeito à combinação linear das variáveis utilizadas na equação de participação. Mais precisamente, os salários de reserva das mulheres seriam positivamente correlacionados com a referida combinação linear.

Por outro lado, uma vez que consideremos as especificações com a correção para seletividade amostral, fica claro o viés para cima no retorno para educação caso se omita o controle para a qualidade. De fato, este resultado é consistente com a análise teórica e resultados empíricos de Behrman e Birdsall (1983), embora no presente caso os diferenciais sejam modestos.

4 COMENTÁRIOS FINAIS

Este artigo pretendeu investigar os retornos para educação a partir da PPV, que é uma base de dados pouco explorada nesse contexto. A análise econométrica preocupou-se especialmente com a questão da seletividade amostral. Assim procedendo, foi possível identificar retornos importantes para a escolaridade formal. Também ficou evidenciada, empiricamente, a importância do controle para a qualidade da educação.

Do ponto de vista metodológico, pelo menos duas questões parecem relevantes. Em primeiro lugar, a consideração de um modelo discreto multinomial no primeiro estágio, quando da estimação da participação no mercado de trabalho, é relevante. Todavia, o modelo Probit multinomial, que seria menos restritivo que o modelo Logit multinomial, adiciona importantes complexidades na estimação. Mais precisamente, o procedimento à Heckman para lidar com seletividade amostral

não está desenvolvido no caso do Probit multinomial e, portanto, pode se constituir em uma linha de pesquisa metodológica a ser explorada.

O segundo ponto refere-se ao fato de que a totalidade dos estudos econométricos acerca de retornos para educação ignora o desenho amostral complexo de pesquisas como a PPV e a PNAD e não considera de forma adequada o erro amostral. Um ramo da literatura estatística pretende dar conta de problemas referentes à amostragem complexa envolvendo mecanismos de estratificação (ver, por exemplo, Pessoa e Nascimento e Silva, 1998). A adaptação de estimadores desta natureza, de forma a levar em conta o problema da seletividade amostral e as propriedades estatísticas dos procedimentos assim definidos, é questão em aberto que merece mais investigações e que pode ter implicações importantes sobre a magnitude efetiva dos retornos para educação. Tais extensões, todavia, extrapolam o escopo do presente trabalho.

Finalmente, fica evidente que uma melhoria nas bases de dados seria desejável para se poder avançar na pesquisa sobre retornos para educação.

REFERÊNCIAS

- Ashenfelter, O.; Harmon, C.; Osterbeek, H. A review of estimates of the schooling/earnings relationship with tests of publication bias. *Labour Economics*, 6, p. 453-470, 1999.
- Barros, R. P.; Corseuil, C. H.; Mendonça, R. Uma análise da estrutura salarial baseada na PPV. *Texto para Discussão n° 689*, IPEA, 1999.
- Berndt, E. R. *The practice of econometrics: classic and contemporary*. New York: Addison-Wesley Publishing Company, 1991.
- Becker, G. S. *Human capital: a theoretical and empirical analysis*. New York: Columbia University Press, 1975.
- Behrman, J.; Birdsall, N. The quality of schooling: quantity alone is misleading. *American Economic Review*, 73, p. 928-946, 1983.
- Behrman, J.; Birdsall, N.; Kaplan R. The quality of schooling and labor market outcomes. In: Sabit, R. H. (ed.), *Opportunities foregone*. Washington-D.C.: IDB, 1996.
- Cairns, G. G. The economic analysis of labor market discrimination: a survey. In: Ashenfelter, O.; Layard, R. (eds.), *Handbook of labor economics*. Amsterdam: North-Holland, 1986.
- Card, D.; Krueger, A. B. Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy*, 100, p. 1-40, 1992.
- Corseuil, C. H. *Estrutura salarial – aspectos conceituais e novos resultados para o Brasil*. Rio de Janeiro: IPEA, 2002.
- Denison, E. F. *Accounting for United States economic growth, 1929-1969*. Washington: Brookings Institution, 1974.
- Ferreira, F. H. G.; Lanjow, P.; Neri, M. A new poverty profile for Brazil using PPV, PNAD and census data. *Texto para Discussão n° 418*, PUC/RJ, 2000.
- Garen, J. The returns to schooling: a selectivity bias approach with a continuous choice variable. *Econometrica*, 52, p. 1199-1218, 1984.
- Greene, W. *Econometric analysis*. New York: Macmillan, 1993.
- Heckman, J. J. Shadow prices, market wages, and labor supply. *Econometrica*, p. 679-694, 1974.
- _____. Sample selection bias as a specification error. *Econometrica*, 47, p. 153-162, 1979.
- Heckman J. J.; Lochner, L.; Todd, P. Fifty years of Mincer earnings regressions. *Discussion Paper, n° 775*, Institute for the Study of Labor (IZA), 2003.

- Kassouf, A. L. The wage rate estimation using the Heckman procedure. *Revista de Econometria*, p. 89-107, 1994.
- _____. Wage gender discrimination and segmentation in the Brazilian labor market. *Economia Aplicada*, v. 2, n. 2, p. 243-269, abr./jun. 1998.
- Lam, D.; Levison, D. Idade, experiência, escolaridade e diferenciais de renda: Estados Unidos e Brasil. *Pesquisa e Planejamento Econômico*, p. 219-256, 1990.
- Lam, D.; Schoeni, R. Effects of family background on earnings and returns to schooling: evidence from Brazil. *Journal of Political Economy*, 101, p. 710-739, 1993.
- Leal, C. I. S.; Werlang, S. R. C. Retornos em educação no Brasil: 1976/89. *Pesquisa e Planejamento Econômico*, p. 559-574, 1991.
- Loureiro, P. R. A.; Galvão, F. G. Discriminação no mercado de trabalho: uma análise dos setores rural e urbano no Brasil. *Economia Aplicada*, v. 5, n. 3, p. 519-545, jul./set. 2001.
- Menezes-Filho, N.; Picchetti, P.; Fernandes, R. A evolução da distribuição dos salários no Brasil: fatos estilizados para as décadas de 80 e 90. In: *Desigualdade e pobreza no Brasil*. Rio de Janeiro: IPEA, 2000.
- Mincer, J. Investment in human capital and personal income distribution. *Journal of Political Economy*, 66, p. 281-302, 1958.
- _____. On-the-job training, costs, returns, and some implications. *Journal of Political Economy*, p. 50-79, 1962.
- _____. *Schooling, experience and earnings*. New York: Columbia University Press, 1974.
- Pessoa, D. G. C.; Nascimento e Silva, P. L. *Análise de dados amostrais complexos*. 13º SINAPE, Caxambu: Associação Brasileira de Estatística, 1998.
- Psacharopoulos, G. Returns to education: an updated international comparison. *Comparative Education*, p. 321-341, 1981.
- Sachsida, A.; Loureiro, P. R. A.; Mendonça, M. J. C. Um estudo sobre retornos em escolaridade no Brasil. *Revista Brasileira de Economia*, 58, p. 249-265, 2004.
- Sedlacek, G. L.; Santos, E. C. A mulher cônjuge no mercado de trabalho como estratégia de geração de renda familiar. *Pesquisa e Planejamento Econômico*, 21, p. 449-470, 1991.
- Soares, R. R.; Gonzaga, G. Determinação de salários no Brasil: dualidade ou não-linearidade no retorno à educação. *Revista de Econometria*, v. 19, n. 2, p. 377-404, 1999.
- Ueda, E. M.; Hoffmann, R. Estimando o retorno em educação no Brasil. *Economia Aplicada*, v. 6, n. 2, p. 209-238, abr./jun. 2002.
- Willis, R. J. Wage determinants: a survey and reinterpretation of human capital earnings functions. In: Ashenfelter, O.; Layard, R. (eds.), *Handbook of labor economics*. Amsterdam: North-Holland, 1986, p. 562-602.

APÊNDICE

Estudos econométricos recentes sobre retornos para educação no Brasil

Estudo	Fonte de Dados	Metodologia	Retornos em Educação
Behrman, J.; Birdsall, N. (1983), <i>The Quality of Schooling: Quantity Alone is Misleading</i> , <i>American Economic Review</i> , 73, p. 928-946	Censo Demográfico 1970	Estimação do retorno via equação de Mincer, modificada para considerar a qualidade da educação. A qualidade é definida conforme a situação geográfica do domicílio em função da alocação de recursos públicos destinados à educação, expressa pela qualificação do professor.	Versão quadrática para o efeito da qualidade gera vários modelos com estimativas entre 11,1% e 20,5% (somente para homens).
Lam, D.; Levison, D. (1990), <i>Idade, experiência, escolaridade e diferenciais de renda: Estados Unidos e Brasil</i> . <i>Pesquisa e Planejamento Econômico</i> , p. 219-256	PNAB/IBGE 1985	Realiza uma decomposição de perfis, para comparar homens brasileiros e norte-americanos, usando o modelo de capital humano. Equações de retorno são estimadas para grupos de idade, experiência, regiões e situação do domicílio.	Retornos para o Brasil: - Por idade: entre 9,4% e 16,4% (> de 18 anos) - Por experiência: entre 14,4% e 20,6% - Por regiões: Sudeste 8,7% a 15% e Nordeste 0,2% a 15,7%
Leal, C.I.S.; Werlang, S.R.C. (1991), <i>Retornos em educação no Brasil: 1976/89</i> . <i>Pesquisa e Planejamento Econômico</i> , p. 559-574	PNAD/IBGE 1976-1989 e Censo Demográfico de 1980	Estimativas para a taxa de retorno pessoal, empregando a equação de Mincer, tendo como explicativas as variáveis experiência e experiência ao quadrado e um conjunto de dummies para vários níveis de escolaridade.	Retornos elevados para as quatro séries iniciais, superior e secundário (16%). Menor para as quatro séries finais do primário (10%).
Sedlacek, G.L.; Santos, E.C. (1991), <i>A mulher cônjuge no mercado de trabalho como estratégia de geração de renda familiar</i> . <i>Pesquisa e Planejamento Econômico</i> , 21, p. 449-470	PNAD/IBGE 1983-1988	Estudo sobre a participação da mulher no mercado de trabalho, específico para cônjuges. Controle para a renda dos maridos e equações de probit para as áreas metropolitanas de RJ, SP e Recife. Uso de dummies para a escolaridade em três níveis.	Coefficientes da escolaridade estimados negativos para o nível mais baixo (exceto RJ) e positivos para além de 8 anos de estudo.
Lam, D., Schoeni, R. (1993) <i>Effects of family background on earnings and returns to schooling: evidence from Brazil</i> . <i>Journal of Political Economy</i> , 101, p. 710-739	PNAD/IBGE 1982	O estudo procura incorporar à equação de retornos variáveis relativas à estrutura e características das famílias dos indivíduos. Escolaridade dos pais e da esposa, situação do domicílio e raça são considerados. Os dados se restringem a homens casados e os retornos são supostos não-lineares.	As estimativas para as equações sem e com as variáveis familiares, por anos de estudo são: 1 ano - 23,4% e 18,2% 8 anos - 140% e 98% 15 anos - 246% e 168%
Kassouf, A.L. (1994), <i>The wage rate estimation using the Heckman procedure</i> , <i>Revista de Econometria</i> , p. 89-107	Pesquisa Nacional de Saúde e Nutrição. IBGE, IPEA e INAN - 1989	Procedimento de Heckman em dois estágios com o emprego da razão de Mills para evitar erro de especificação (heteroscedasticidade tratada com matriz de White). Estimativas por gênero.	Taxa de retorno no caso feminino (8,5%) pouco superior à masculina (7,2%). O contrário sem Heckman e Mills (7,0% contra 7,7%)

Estudo	Fonte de Dados	Metodologia	Retornos em Educação
Birdsall, N., Behrman, J. (1996), <i>The Quality of Schooling and Labor Market Outcomes</i> , In R.H. Sabit (ed.), <i>Opportunities Foregone</i> , Washington-D.C.: IDB	Censos Demográficos 1970/80	A qualidade da educação e o efeito de fatores específicos (grupos de indivíduos) são incorporados à equação de rendimentos. A qualidade é avaliada pela escolaridade dos professores e outros controles são feitos com dados de homens em grupos de idade distintos para os levantamentos dos censos de 1970 e 1980.	Conforme o conjunto de explicativas os retornos foram: - Homens de 25 a 45 anos de idade: Entre 17,4% e 29,6% - Homens de 15 a 35 anos de idade: Entre 15,4% e 41,8%.
Kassouf, A.L. (1998), <i>Wage gender discrimination and segmentation in the Brazilian labor market</i> , <i>Economia Aplicada</i> , p. 243-269	Pesquisa Nacional de Saúde e Nutrição. IBGE, IPEA e INAN - 1989	No primeiro estágio do procedimento de Heckman utiliza-se, ao invés da versão binária, um modelo multinomial LOGIT tendo em vista a perspectiva de estratégias distintas para os ocupados no setor formal, informal e os não ocupados. São realizadas estimativas por gêneros, sendo controladas por raças e regiões do país.	Taxas de Retornos em educação: Caso masculino: - Formal 19,1% - Informal 6,2% Caso feminino: - Formal 21,9% - Informal 14,5%
Soares, R.R. , Gonzaga, G. (1999) <i>Determinação de salários no Brasil: dualidade ou não-linearidade no retorno à educação</i> . <i>Revista de Econometria</i> , 19(2):377-404	PNAD/IBGE 1988	Determinação endógena dos setores duais por meio do modelo switching-regressions proposta por Dickens e Lang (1985). Estimação de modelos do tipo dual, competitivo e unissetor não linear.	-Dual: Primário 23,4% e secundário 11,8%. -Competitivo: 15,4% -Não Linear: De -0,31% até 84,4%.
Menezes-Filho, N., Picchetti, P., Fernandes, R. (2000), <i>A evolução da distribuição dos salários no Brasil: fatos estilizados para as décadas de 80 e 90</i> . <i>Desigualdade e Pobreza no Brasil</i> . Rio de Janeiro: IPEA.	PNAD/IBGE 1977-1997	Por meio da formulação de MaCurdy e Mroz (salário em função do tempo da idade e do grupamento), aplicada para repetidas cross-section estima-se os retornos para educação lineares com os dados grupados da amostra para quantis de renda e níveis educacionais, buscando evidência empírica para as motivações da desigualdade salarial no BR.	A trajetória no tempo o s retornos depende do nível de escolaridade. A diferença dos grupos de > 11 anos com relação ao 9-11 está entre 7,5% a 9,6%, crescente. Nos outros casos caem (de cerca de 6,5% a 5,0%).
Loureiro, P.R.A., Galvão, F.G. (2001). <i>Discriminação no Mercado de Trabalho: Uma análise dos setores rural e urbano no Brasil</i> . <i>Economia Aplicada</i> , v. 5, n. 3, p. 519-545.	PNAD/IBGE 1998	Decomposição de causas por meio de Blinder-Oaxara (equações distintas para cada valor dos atributos incidentes), combinada ao procedimento de dois estágios de Heckman. Estimação por raças, gêneros, situação do domicílio.	Retorno em educação: -Urbano: homens 16,8%, mulheres 20,9%, pardos 14,7% e brancos 11,4%; - Rural: homens 12,7%, mulheres 16,3%, Pardos 11,6% e Brancos 12,5%;

Estudo	Fonte de Dados	Metodologia	Retornos em Educação
Ueda, E.M., Hoffmann, R. (2002). Estimando o retorno em educação no Brasil. <i>Economia Aplicada</i> , v. 6, n. 2, p. 209-238	PNAD/IBGE 1996	É conduzida uma análise sobre a fragilidade de OLS para retornos em educação e duas alternativas econométricas são estudadas. São discutidos MVI e métodos que buscam controlar características individuais. Neste sentido, duas alternativas: um mesmo indivíduo ao longo do tempo ou indivíduos diferentes com indicadores intra-familiares próximos (características dos pais).	Dois modelos com a variável educação sendo discreta ou contínua e 2 técnicas: MQO e MVI. -COM controle por país: Entre 9,3% e 136,8%. -SEM controle por país: Entre 11,2% e 166,2%
Sachsida, A., Loureiro, P.R.A., Mendonça, M.J.C. (2004), Um Estudo sobre Retornos em Escolaridade no Brasil, <i>Revista Brasileira de Economia</i> , 58, 249-265	PNAD/IBGE 1992/1999	Procedimentos em cross-section e dados empilhados. Quatro modelos estimados com conjuntos de explicativas distintos. Técnicas de OLS, Heckman, Garen e Pseudopainel.	Estimativas que variam, usando o procedimento de Heckman, de 12,9%, em cross-section, até 14,7%, para o caso de dados empilhados.