

**Endogeneidade da Educação na Previsão da Taxa de Retorno: Avaliação Metodológica e  
Aplicação para Regiões Brasileiras e Estados Selecionados**

Francisca Zilania Mariano  
Mestre, Doutoranda em Economia, CAEN/UFC  
[ainaliz@yahoo.com.br](mailto:ainaliz@yahoo.com.br)

Ronaldo A. Arraes  
PhD, Professor CAEN/UFC  
[ronald@ufc.br](mailto:ronald@ufc.br)

Endereço: Av. da Universidade, 2700  
60020-181 Fortaleza – CE

Área de interesse: 2 – Economia Social

## **Endogeneidade da Educação na Previsão da Taxa de Retorno: Avaliação Metodológica e Aplicação para Regiões Brasileiras e Estados Selecionados**

### **Resumo**

Tendo em vista o problema da endogeneidade da variável educação e as fontes de dados disponíveis para o Brasil, comparam-se dois métodos de estimação para a equação minceriana de rendimentos, visando inferir sobre qual deles possui o melhor poder preditivo, modelo endógeno ou o procedimento com variáveis instrumentais (VI). Verifica-se que o método de Heckman aplicado ao primeiro modelo possui o menor erro quadrático médio (MSE). Isso acontece devido à qualidade dos instrumentos disponíveis, o que alerta para a necessidade de haver uma maior interação entre as agências responsáveis pelo desenho e coleta de dados e a comunidade científica sobre as informações a serem geradas em uma *survey* para atender as ações de políticas locais com diagnósticos mais precisos. Constatado que o modelo endógeno apresenta melhor poder preditivo, estimam-se variações na equação de rendimentos através da interação entre capital humano e regionalização. Observa-se que as taxas de retorno da educação variam de acordo com as regiões e estados brasileiros, sendo a região Nordeste e o estado do Ceará os que apresentam o menor retorno, equivalente a, aproximadamente 14,68%, e ao compará-los a região Centro-Oeste e ao estado de Santa Catarina o diferencial entre os retornos chega a 26% e a 31%, respectivamente. Uma vez corrigido parte do viés de habilidade, isso mostra que indivíduos advindos de regiões com distinto grau de desenvolvimento econômico não acumulam igualmente o mesmo nível de capital humano.

Palavras chave: Endogeneidade e Retorno Educação, Equação Minceriana, Modelos Preditivos

### **Abstract**

In view of the endogeneity problem in the education variable and the available data sources for Brazil, two estimation methods for the mincerian earning equation are compared to infer about which one has the best predictive power – endogenous model or the procedure with instrumental variables (IV). It is found that the method of Heckman applied to the first model has the lowest mean squared error (MSE). This happens due to the quality of IV, which alert to the need for greater interaction between the agencies responsible for the design and data collection and the scientific community about the information to be generated in a survey to meet local political actions with diagnoses more accurate. Furthermore, given better predictive power of the endogenous model, alternative specifications in the earning equations are performed in order to allowing average returns to be variable as human capital and regionalization interact. It is observed that the education rates of return of education vary according to regions and Brazilian States, where the Northeast region and the State of Ceará present the lowest returns, equivalent to approximately 14.68%, which compared to the Mid-West region and the State of Santa Catarina the differential between the returns reaches 26% and 31%, respectively. This corrects part of the skill bias, and shows that individuals from regions with distinct degree of economic development do not accumulate equally the same level of human capital.

Keywords: Endogeneity and Return to Education, Mincerian Equation, Predictive Models

## 1. Introdução

A relação entre escolaridade e rendimentos é um dos mais frequentes assuntos abordado em economia empírica. Um elevado número destes estudos desenvolve versões da equação de salário proposta por Mincer (1974), cujo parâmetro-chave está associado com anos de escolaridade, derivado para capturar os efeitos sobre as diferenças de rendimentos causadas por diferenças de escolaridade formal dos indivíduos. Entretanto, para dar uma interpretação causal dos parâmetros nessa equação, deve-se ter em conta que a variável explicativa “anos de escolaridade” é endógena, vez que é intrinsecamente definida como uma variável de escolha. O problema de endogeneidade está relacionado ao fato de que o pesquisador não observa todos os factores que afetam a escolha desejada de escolaridade. Se alguns desses fatores não observados estão correlacionados com os observados na equação de rendimentos, então, estimações convencionais produziriam estimativas visadas do retorno da escolaridade (denominado viés de habilidade).

Estimativas da taxa de retorno da educação variam largamente, dependendo da base de dados utilizada, das hipóteses assumidas e das técnicas de estimação utilizadas. Em termos metodológicos, o debate sobre a endogeneidade requer, em geral, identificar hipóteses que não podem ser empiricamente testadas ou se mostram fracas em processos de estimação. Além disso, ao se proceder estimativa de uma única taxa de retorno pode não fornecer informação relevante se retornos da educação diferem por nível de escolaridade, ou entre populações (estratos sociais) ou entre regiões heterogêneas. Afinal, sua importância para respostas de políticas são contrapostas ao debate metodológico. Nesse sentido, o risco desempenha um importante papel na decisão de investimento na educação do indivíduo, bem como no nível de investimento educacional do Governo, que devem ser levados em consideração na tomada de decisão sobre racionalidade e otimização do investimento em educação (HECKMAN, LECHNER, TODD, 2008).

Teoricamente presume-se que o viés de habilidade ocorra devido à correlação entre duração dedicada à educação e o termo de erro na equação de rendimentos. Se há tal correlação e é positiva, como se espera que exista, isso implica que os indivíduos com elevado retorno do trabalho (independentemente do nível de escolaridade) sistematicamente escolhem um nível de escolaridade maior do aqueles em situação oposta, resultando, daí, a questão de heterogeneidade que deve ser tratada na especificação teórica do modelo e o correto tratamento econométrico para lidar com esse problema. Vários estudos econométricos recentes têm levado em conta que há ocorrência de heterogeneidade não apenas associada com a capacidade de rendimento em geral, mas também de heterogeneidade associado ao retorno da escolaridade, ou seja, alguns indivíduos ganham mais por um ano adicional de escolaridade do que outros. Willis e Rosen (1979) abordaram essa questão em seu modelo teórico. Heterogeneidade deste tipo é frequentemente denominada vantagem comparativa, e é normalmente tratada através da formulação de um modelo com coeficiente aleatório, no qual o coeficiente associado com anos de escolaridade pode variar entre indivíduos de acordo com alguma função de distribuição. Se este coeficiente aleatório estiver correlacionado com a variável de escolaridade ou o termo de erro na equação de rendimentos, então estimativas padrão de mínimos quadrados dos retornos da escolaridade são viesados.

Para tratar com o viés de habilidade e endogeneidade da escolaridade, técnicas de variável instrumental têm sido comumente aplicadas. Como resultado, tem surgido uma vasta literatura sobre como interpretar estimativas de variável instrumental no caso de heterogeneidade do retorno da educação (WOOLDRIDGE, 2002; HECKMAN E VYTLACIL, 2005). Relacionado a isso está o procedimento em dois estágios, no qual, no primeiro estágio, estima-se uma equação de escolaridade com a inserção de dequadas

variáveis exógenas que servem como instrumentos na equação de rendimentos em um segundo estágio visando a eliminação do viés (HECKMAN, 1979; CARD, 2001).

Além do foco sobre os diferentes tipos de vieses de seleção, atenção crescente tem sido dada para a especificação da equação de Mincer na literatura. Uma das características teóricas dessa equação é a relação linear entre o logaritmo dos rendimentos e escolaridade, enquanto outra é a suposta separabilidade entre escolaridade e experiência. Alguns autores como Heckman e Polachek (1974), Heckman, Lochner e Todd (2008) e Belzil (2007) analisaram sua validade, e as consequências em relaxá-la, e as suposições de outras formas funcionais do esquema padrão minceriano. O resultado geral dos testes dessas investigações é que algumas hipóteses simplificadoras são rejeitadas e, portanto, que há margem para que o modelo se torne mais flexível.

Nesse sentido, o problema central desta pesquisa se apóia nesses fatos como questões norteadoras básicas e propõe investigar, levando em consideração a endogeneidade da educação e as fontes de dados disponíveis para o Brasil, se o método para correção deste problema através da aplicação de variáveis instrumentais fornece estimativas mais confiáveis. Sendo o capital humano, o principal determinante das variações nos rendimentos, cabe distinguir, em contribuição do estudo, seus efeitos quando comparados indivíduos provenientes de regiões com diferentes estágios de desenvolvimento econômico. Isso implica em permitir retornos variáveis do capital humano.

Em consonância com o problema e com a literatura, o objetivo geral deste trabalho é o de analisar o problema da endogeneidade e retornos da educação no Brasil. Com isso, este trabalho pretende, inicialmente, comparar dois métodos de estimação para a equação de rendimentos, modelo endógeno e o procedimento de variáveis instrumentais, considerando a amostra para o Brasil, as regiões Nordeste e Sul-Sudeste, e para os estados do Ceará e Rio de Janeiro, para se inferir sobre qual desses possui o melhor poder preditivo. Após a estimação dos dois métodos, será escolhido o que apresenta o menor erro quadrático médio (MSE) para se estimar variações na equação de rendimentos e verificar os diferenciais do efeito do capital humano, representado pela escolaridade, sobre as regiões Nordeste, Sul, Sudeste e Centro Oeste, e estados selecionados de cada uma destas, Ceará, Santa Catarina, São Paulo e Mato Grosso do Sul. Esse procedimento certamente reduz o viés de endogeneidade causado tanto pela escolha educacional como o viés de habilidades dos indivíduos. Além disso, busca-se mostrar se o retorno da educação varia de acordo com a faixa de escolaridade em que o indivíduo se encontra, testando, assim, a hipótese do modelo teórico de Mincer (1974) que o retorno futuro esperado independe do nível de escolaridade escolhido pelo indivíduo. Além disso, no processo de estimação da equação minceriana incorpora-se o problema de seletividade amostral, cuja correção elimina a possibilidade de um viés adicional nas estimativas. Dados da Pesquisa Nacional por Amostras de Domicílios (PNAD) da Fundação Instituto Brasileiro de Geografia e Estatística (FIBGE) para o ano de 2009 dão suporte empírico ao estudo.

O trabalho está organizado de modo a expor, em sequência, a revisão de literatura sobre capital humano e retornos da educação, seguida pela metodologia, resultados e conclusões.

## 2. Capital Humano e Retornos da Educação

Não obstante a elegância matemática desenvolvida e importância reconhecida para a literatura no artigo seminal de Mincer (1958), algumas hipóteses nele contidas, como a uniformidade e indiferença entre os indivíduos acerca dos níveis de escolaridade, bem como a experiência, geraria indeterminação na escolha de escolaridade ao nível individual. Os refinamentos de modelagem incorporados em Mincer (1974), extraídos das análises de Becker e Chriswick (1966) e Ben-Porath (1967), permitiram incorporar experiência do trabalho pós-educação formal como determinante dos rendimentos e aproximar um modelo de otimização com escolha de escolaridade e experiência do trabalho, assumindo que a proporção dos rendimentos abdicada em investimento declinasse com experiência. Esse esquema de análise permitiu a incorporação de heterogeneidade entre os agentes, resultando na denominada equação minceriana de rendimento, vastamente utilizada em aplicações empíricas na literatura mundial.

A maioria dos trabalhos aplicados se baseia nas propostas destes autores para estimar equações de rendimentos, os quais, tais como as formulações originais, especificam a educação como única variável representativa do capital humano, e o coeficiente que mede a taxa de retorno é idêntico entre os indivíduos. Com isso, por ser a principal variável que dita as variações nos rendimentos, a escolha de uma medida para capital humano é um problema secular em estudos empíricos nos países com notória deficiência na formação de bancos de dados. Por isso a escolha única recai na escolaridade para representá-la. Vale ressaltar que, em tese, esta *proxy* não seria a ideal, pois, segundo a teoria do capital humano este refere-se ao conjunto de habilidades e capacidades do indivíduo que afeta a sua produtividade e depende dos investimentos realizados ao longo do ciclo de vida, tais como em educação, treinamento profissional, hábitos de vida saudável, aquisição de bens e serviços de saúde.

Tornou-se senso comum afirmar que a educação é a variável mais importante na determinação da desigualdade de renda. Com o objetivo de explicar a desigualdade para os Estados Unidos, utilizando dados do censo de 1960, Mincer (1974) desenvolve uma equação de rendimentos utilizando somente o nível educacional como representativo do capital humano e constata que um ano adicional de escolaridade gera um aumento de 11,5% nos rendimentos.

Seguindo a proposta de Mincer, vários autores utilizam essa equação para calcular os retornos da educação sobre os ganhos anuais em diferentes países. Fleischhauer (2007) ao analisar diversos trabalhos que utilizaram esse procedimento, tais como: Ashenfelter e Krueger (1994), Rouse (1999), Arias e McMahon (2001), os quais afirmaram que os retornos da educação variam de 12% a 16%, 10%, 13% e de 5% a 10%, respectivamente, conclui que esses resultados geralmente resultam em estimativas entre 5% e 15%.

Embora Boudarbat, Lemieux e Riddell (2008) reconheçam que o estoque de capital humano acumulado por um indivíduo é resultado de uma combinação de vários fatores, os autores utilizam somente o nível educacional e a experiência de trabalho para analisar a evolução dos retornos do capital humano no Canadá sobre as diferenças salariais, no período de 1980 a 2006. Observaram que os retornos à educação aumentaram substancialmente para os homens canadenses, entre 1980 e 2000, passando de 34% para 43% e para as mulheres. Para o período mais recente de 2000-2006 os resultados mostram uma tendência decrescente das diferenças salariais medidos pelo retorno à educação.

Zepeda et al. (2009) identificam os fatores que explicam as alterações observadas na média dos rendimentos em três países da América Latina – Brasil, Chile e México – utilizando variáveis demográficas e socioeconômicas, dentre elas, o nível educacional como representativo para o capital humano, e constatam que a única mudança mais significativa em

todos os países e períodos, foi o impacto do nível educacional sobre a força de trabalho, indicando que a escolaridade é o fator mais importante para o aumento nos rendimentos. De forma mais criteriosa e ampla, Sapelli (2009), ao verificar as taxas de retornos da educação no Chile, utilizando uma metodologia baseada no modelo de Mincer (1974), afirma que estas taxas irão variar dependendo do nível educacional, divididos em três grupos, educação básica, média e superior, conforme comprovado com os resultados empíricos de 5,6%, 14,3% e 25%, respectivamente. Além disso, o trabalho compara estes resultados com os observados em outros países – Argentina, Colômbia, México e Brasil – e conclui que apenas os retornos no ensino básico se equiparam com os do Chile, perdendo nos demais níveis educacionais.

No Brasil, também existem problemas de coleta, acesso, qualidade dos dados e escolha de *proxies* eficientes na estimação dos modelos que possam estabelecer uma melhor conexão entre o vetor de variáveis que compõem o capital humano e as informações disponíveis. Por isso, a maioria dos trabalhos empíricos também utiliza somente o nível educacional para explicar a desigualdade de renda, tais como: Langoni (1973), Lam e Levinson (1992), Soares e Gonzaga (1999), Menezes Filho (2001), Ueda e Hoffman (2002), Budría e Pereira (2005), Ramos (2007), Barros, Franco e Mendonça (2007).

Ao apresentar um dos trabalhos pioneiros sobre esse tema no Brasil, Langoni (1973) discorre sobre o processo de geração das desigualdades de renda, ou seja, a forma como tais desigualdades são geradas e reveladas no mercado de trabalho. Sua investigação mostra que o nível educacional, idade, gênero, setor de atividade e região de residência são determinantes dos diferenciais salariais.

Quanto aos retornos da educação no Brasil, Psacharopoulos (1987), Lam e Levinson (1992) constataram ser o salário pela educação um dos mais altos do mundo, com uma taxa em torno de 15%. Já Loureiro e Galvão (2001) estimaram o retorno da educação na zona rural e urbana e encontram taxas de 11% e 18%, respectivamente. Soares e Gonzaga (1999) utilizaram um modelo *switching-regressions* e encontraram um retorno marginal de 15,4 referente a um ano a mais de estudo. Sachsida, Loureiro e Mendonça (2004) estimam o retorno da educação, utilizando os procedimentos de Heckman (1979), em cross-section, pooling regressions e pseudopanel. Os autores encontram retornos que variam de 10% a 22%, dependendo do método e do nível de escolaridade utilizado. Também de acordo, Resende e Wyllie (2005) investigaram os retornos para a educação utilizando como base de dados a Pesquisa sobre Padrão de Vida (PPV-IBGE) para 1996 e 1997 e encontraram taxas de 12,6% e 15,9% para mulheres e homens, respectivamente.

É consenso na literatura a influência da educação para ditar a queda da desigualdade de renda no Brasil, via aumento de produtividade dos trabalhadores menos qualificados. Em decorrência, observa-se empiricamente, em níveis de países ou regiões, heterogeneidade das taxas de retorno da educação, as quais decrescem com os seus graus de desenvolvimento, dando suporte a afirmação de Lopez, Thomas e Wang (1998) apud Bezerra e Ramos (2008) que a produtividade marginal do capital humano difere espacialmente. Com isso, do ponto de vista teórico seria inaceitável ter-se somente uma taxa de retorno da educação para representar o Brasil, principalmente devido a existência de regiões com distintos estágios de desenvolvimento. No caso em estudo, não se espera que indivíduos com mesmo número de anos de estudo nas regiões Nordeste e Sul-Sudeste tenham acumulado igualmente o mesmo nível de capital humano, ainda que considerando apenas o aprendizado cognitivo, conseqüentemente, deve-se permitir variação de seu efeito marginal sobre os rendimentos.

Bezerra e Ramos (2008) analisam a distribuição do capital humano no Brasil para o período de 1981 a 2002, concentrando na região Nordeste, pois esta apresenta o menor produto per capita entre as regiões. Segundo os autores é importante saber como a educação se distribui entre as regiões para poder explicar as divergências regionais quanto ao

crescimento do produto e quanto à distribuição de renda. Concluem que a distribuição da educação difere entre as regiões do Brasil, pois um ano adicional de estudo na região Sul e na região Nordeste têm impactos diferentes sobre o produto de cada região, sendo, portanto, importante saber a forma como esse capital se distribui. Além disso, afirmam que embora tenha havido uma redução na desigualdade educacional no período analisado, as diferenças entre as regiões ainda persistem, pois não houve significativa redistribuição de capital humano entre as mesmas, com a região Nordeste liderando a posição de pior distribuição educacional.

### 3. Metodologia

#### 3.1 Base de Dados

A verificação empírica deste estudo está fundamentada na base de dados da PNAD/FIBGE sobre informações individuais de características socioeconômicas e regionais para o ano de 2009. Vale ressaltar que foram realizados alguns filtros na amostra, tais como: foram excluídos os militares, pois seus rendimentos não variam de acordo com as “forças” de mercado, além disso, partindo do pressuposto de que estudo e trabalho são mutuamente exclusivos, excluíram-se os indivíduos que estavam estudando e trabalhando e foram selecionados apenas pessoas com idade entre 15 e 35 anos.

#### 3.2 Equações Mincerianas

A formulação econométrica do presente estudo visa, essencialmente, quantificar de forma teoricamente mais apropriada o efeito do capital humano sobre o rendimento e, conseqüentemente, sobre os diferenciais de produtividade que geram as divergências advindas de heterogeneidade do retorno desse capital sobre os rendimentos. Embora as discussões precursoras acerca da magnitude da taxa de retorno de educação na renda dos indivíduos devam-se a Schultz (1961) e Ben-Porath (1967), elas foram intensificadas a partir da publicação do artigo seminal de Mincer (1974), o qual formalizou teoricamente a derivação da equação de rendimentos tendo como variável dependente a renda sendo explicada por educação e experiência. Com base em sua formulação inicial de otimização do valor presente dos rendimentos futuros, ele estabeleceu um modelo de equilíbrio para determinação dos rendimentos individuais de onde extraiu uma forma para se estimar a taxa de retorno da escolaridade. Embora derivada em hipóteses rígidas, como exemplo de indivíduos possuírem idênticas habilidades e desconsiderar experiência (investimento) pós-educação, ele se baseou nas análises de Becker e Chriswick (1966) e Ben-Porath (1967) para estabelecer a segunda versão de seu modelo, no qual assume que a proporção da renda abdicada por investimentos pós-educação é a mesma em todos os níveis de escolaridade. Com base nisso, diferenciam-se as habilidades dos indivíduos diante suas escolhas e o resultado do retorno é em termos médios, conforme a equação de Mincer (1974):

$$\ln w(s, x) = \alpha_0 + \rho s + \beta_0 x + \beta_1 x^2 + \epsilon \quad (1)$$

Onde  $\rho$  é a taxa média de retorno da escolaridade ( $s$ ) entre os indivíduos, e  $x$  é a experiência adquirida pós-educação. A originalidade de sua formulação inicial (1958) e a equação de estimação ajustada em sua forma simplificada (1974) são responsáveis pela popularidade desta última como estratégia adequada para se estimar taxa de retorno da educação, conforme Heckman et al (2008).

Cabe reprimir que a maioria dos trabalhos sobre os elementos determinantes dos rendimentos é fundamentada no modelo proposto por Mincer (1974), sob a hipótese fundamental de que a escolaridade de um indivíduo – medição usual do capital humano – é o principal elemento condutor das diferenças salariais dos trabalhadores (LANGONI, 1973;

MINCER, 1974; LAM; LEVINSON, 1992; MENEZES FILHO, 2001; DUARTE, 2003; HECKMAN, 2005). Esse modelo em sua especificação original relaciona o logaritmo da razão salário-hora aos anos de estudo (E) e anos de experiência (EXP) na forma linear e quadrática (Heckman, 2005). Sua aplicação é bastante difundida na área de economia do trabalho, devido ao fato de o coeficiente estimado da variável educacional (E) poder ser interpretado como a taxa marginal de retorno da educação ou retorno esperado da produtividade do trabalhador (HECKMAN, 2005; FLEISCHHAUER, 2007).

Para fins de comparação e suporte metodológico do modelo a ser aqui proposto, cabe, inicialmente, especificar a equação *Minceriana* da seguinte forma:

$$\ln(Y_i) = \beta_0 + \beta_1 E_i + \beta_2 I_i + \beta_3 I_i^2 + \sum_i X_i + \varepsilon_i \quad (2)$$

A contribuição do presente trabalho consiste, inicialmente, em aplicativos empíricos sobre a comparação teórica entre dois métodos de estimação para a equação de rendimentos (2): método de Heckman e o procedimento de variáveis instrumentais. O primeiro corrige o problema da seletividade amostral, porém não leva em consideração o viés de endogeneidade da educação; já o segundo, além de corrigir o primeiro problema, utiliza instrumentos para eliminar este viés, embora possa surgir outro, dependendo da validade econômica e estatística destes instrumentos. Com isso, cabe contrapor as duas alternativas de estimação para se inferir sobre qual delas possui o melhor poder preditivo: o modelo endógeno, decorrente da endogeneidade da variável educação; modelo VI, dependente da qualidade dos instrumentos? Vale ressaltar que nesta primeira análise, a variável experiência será representada pela idade do indivíduo, pois caso fosse construída da forma tradicional (idade – anos de estudo – 6) essa também seria endógena. Além disso, testou-se os dois modelos considerando a amostra para o Brasil, as regiões Nordeste (NE) e Sul-Sudeste (SSE), e para os estados do Ceará (CE) e Rio de Janeiro (RJ).

Após a estimação dos dois métodos citados acima, escolher-se á o que apresenta o menor erro quadrático médio (MSE) para se estimar variações na equação de rendimentos, a fim de permitir retornos médios variáveis. Tais variações serão inicialmente obtidas através da interação do capital humano e regionalização – grandes regiões e estados selecionados de cada uma destas. Nesse sentido, não se espera que indivíduos com mesmo número de anos de estudo, advindos de regiões menos e mais desenvolvidas, tenham acumulado igualmente o mesmo nível de capital humano, ainda que considerando apenas o aprendizado cognitivo, conseqüentemente, deve-se permitir variação de seu efeito marginal sobre os rendimentos. Esse procedimento certamente reduz o viés de endogeneidade causado tanto pela escolha educacional como o viés de habilidades dos indivíduos. Além disso, busca-se mostrar se o retorno da educação varia de acordo com a faixa de escolaridade em que o indivíduo se encontra, testando, assim, a hipótese do modelo teórico de Mincer (1974) que o retorno futuro esperado independe do nível de escolaridade escolhido pelo indivíduo.

Para tanto, serão estimadas duas equações para captar a heterogeneidade do retorno da educação entre as regiões, fazendo na equação (3)  $j = 1, 2, 3$  e 4, e entre os estados, fazendo  $j = 5, 6, 7$  e 8, representadas da seguinte forma:

$$\ln Y_i = \beta_0 + \beta_1 E_i + \beta_2 E_i * D_{ij} + \beta_3 E_i * D_{ij} + \beta_4 E_i * D_{ij} + \beta_5 E_i * D_{ij} + \beta_6 EXP_i + \beta_7 EXP_i^2 + \beta_8 G_i + \beta_9 R_i + \beta_{10} U_i + \beta_{11} \varphi + \varepsilon_i \quad (3)$$

Em seguida, será estimada a equação (4) definida por faixa de escolaridade, faixas I, II e III, as quais representam os indivíduos que possuem cinco a nove anos de estudo, dez a doze e treze ou mais anos de estudo, respectivamente.

$$\ln Y_i = \beta_0 + \beta_1 E_i + \beta_2 E_i * EXP_i + \beta_3 EXP_i + \beta_4 EXP_i^2 + \beta_5 G_i + \beta_6 R_i + \beta_7 U_i + \beta_8 \varphi + \varepsilon_i \quad (4)$$

As variáveis utilizadas nos modelos estão descritas no quadro a seguir:

<b>Variável Dependente</b>	<b>Descrição</b>
<i>Y</i>	Rendimento do trabalho principal do indivíduo
<b>Variáveis Explicativas</b>	<b>Descrição</b>
<i>E</i>	Anos de Estudo
<i>I</i>	Idade do indivíduo
<i>EXP</i>	Idade – Anos de Estudo – 6
<i>G</i>	Gênero: assume valor 1 para homens
<i>R</i>	Raça: assume o valor 1 para indivíduos brancos
<i>U</i>	Condição de Urbanização: assume valor 1 para indivíduos residente em áreas urbanas
$\varphi$	Razão inversa de Mills do procedimento de Heckman (1979)
<i>D</i> <sub>1</sub>	Assume valor 1 para indivíduos do Nordeste
<i>D</i> <sub>2</sub>	Assume valor 1 para indivíduos do Sul
<i>D</i> <sub>3</sub>	Assume valor 1 para indivíduos do Sudeste
<i>D</i> <sub>4</sub>	Assume valor 1 para indivíduos do Centro-Oeste
<i>D</i> <sub>5</sub>	Assume valor 1 para indivíduos do Ceará
<i>D</i> <sub>6</sub>	Assume valor 1 para indivíduos de Santa Catarina
<i>D</i> <sub>7</sub>	Assume valor 1 para indivíduos de São Paulo
<i>D</i> <sub>8</sub>	Assume valor 1 para indivíduos de Mato Grosso do Sul

Quadro 1 – Variáveis do Modelo

### 3.3 Correções de Fontes de Viés na Equação de Rendimentos

#### 3.3.1 Método de Heckman

Quando se processa a estimação de uma equação de rendimentos utilizando como base de dados pesquisas censitárias existem diferentes fontes de vieses que lhes são inerentes, portanto, não podem ser desconsiderados nesse processo, a exemplo do problema de seletividade amostral. Ao utilizar a PNAD como base dados deve-se levar em consideração que as informações coletadas são fornecidas pelas pessoas que tinham trabalho no período da pesquisa, ou seja, os salários observados nas PNADs estão relacionados com a decisão de um indivíduo trabalhar ou não e esta pode estar correlacionada a fatores não observáveis que afetam seu rendimento. Com isso, a consideração somente de pessoas que estejam trabalhando em uma equação de rendimentos pode produzir estimadores viesados (WOOLDRIDGE, 2006).

Para solucionar o problema do possível viés de seletividade amostral, será aplicado o procedimento de Heckman (1979) em dois estágios, o qual é formado por duas equações; uma que descreve a participação no mercado de trabalho inserindo os indivíduos que auferem renda positiva e os que nada recebem, e outra do tipo minceriana que considera somente os que possuem algum rendimento. O primeiro estágio consiste em estimar a seguinte equação de participação:

$$L = \beta_i X_i + \mu_i \quad (5)$$

Onde  $L$  é a variável que designa a participação no mercado de trabalho e  $X_i$  representa um conjunto de variáveis que explicam a participação no mercado de trabalho. Como  $L$  não é observado, uma variável *dummy* é utilizada:

$$\begin{aligned} L_i &= 1 \text{ se } L_i^* > 0 \\ L_i &= 0 \text{ se } L_i^* \leq 0 \end{aligned}$$

Neste caso,  $L$  assume o valor 1 para representar o indivíduo que está trabalhando e 0 caso contrário. Após a estimação do modelo probit, o segundo passo para aplicar o método de Heckman (1979) consiste em estimar a equação de rendimentos levando em consideração o viés de seleção amostral.

$$W_i = \gamma_i' Z_i + \varepsilon_i \quad (6)$$

Onde  $W$  é o logaritmo da renda do indivíduo,  $Z$  é o vetor de características pessoais e  $\varepsilon$  é o vetor de erros. Assim, o viés de seleção amostral pode ser observado da seguinte forma:

$$\begin{aligned} E[W_i / L_i^* > 0] &= E[W_i / \mu_i > -\beta_i X_i] = \gamma_i' Z_i + E[\varepsilon_i / \mu_i > -\beta_i X_i] \\ &= \gamma_i' Z_i + \frac{\text{cov}(\mu_i, \varepsilon_i) \phi(\beta_i X_i)}{\sigma_\mu \Phi(\beta_i X_i)} \end{aligned}$$

Em que  $\phi$  é a função de densidade de probabilidade e  $\Phi$  é a função de densidade normal padronizada acumulada. A função  $\varphi_i = \frac{\phi(\beta_i X_i)}{\Phi(\beta_i X_i)}$  é denominada razão inversa de Mills e representa a variável que segundo Heckman (1979) é utilizada como explicativa na regressão principal para eliminar o viés de seletividade amostral, onde a significância desta função indicará a presença de viés de seletividade amostral. Logo, a equação minceriana será representada da seguinte forma:

$$\ln Y = \gamma_i' Z_i + \gamma_\varphi \varphi_i + \varepsilon_i \quad (7)$$

$$\ln Y = \gamma_i' K_i + u_i \quad (8)$$

A estimação da equação (6) pelo método de mínimos quadrados ordinários forneceria estimativas inconsistentes, devido a omissão da variável  $\varphi_i$ , resultando no erro de especificação descrito por Heckman (1979).

### 3.3.2 Estimação por Variáveis Instrumentais

Outro tipo de viés encontrado em uma equação de rendimentos é o problema da endogeneidade da educação, pois Cameron e Trivedi (2005, p.92-93, 100-102) mostram que a omissão e/ou a característica endógena de uma variável em uma equação, especificamente de rendimentos, poderá causar um problema grave de viés e inconsistência nos estimadores quando se aplica o usual método de mínimos quadrados ordinários (MQO). Para corrigir esses efeitos danosos no processo de estimação a alternativa teórica é aplicar o método em dois estágios com procedimento de variáveis instrumentais, onde estas devem ser correlacionadas com a variável endógena e exógena ao modelo. O problema que persiste e que a maioria dos trabalhos desconsidera refere-se à qualidade destes instrumentos.

Na literatura são encontrados diversos trabalhos que corrigem o viés da endogeneidade da educação através de VI, porém, divergem sobre as escolhas dos instrumentos, seja por inadequação voluntária ou indisponibilidade de dados. Suliano e

Siqueira (2010) estimam as taxas de retornos da educação para estados da região Nordeste do Brasil, através da PNAD, utilizando como instrumento o número de pessoas na família do indivíduo, com o argumento de que educação e fertilidade são decisões interdependentes, ou seja, dadas as limitações orçamentárias, indivíduos que decidem ter mais filhos passam a investir menos em educação. Já Salvato e Silva (2008) utilizam a renda de aluguel como instrumento na estimação para a região metropolitana de Belo Horizonte, justificando que este é um dos rendimentos de não trabalho e que pode ser entendida como uma proxy para o nível de renda da família de origem, mostrando assim as condições da família em que o indivíduo foi educado.

Para verificar as mudanças nas taxas de retornos da educação na Argentina, Di Pietro e Pedace (2008) utilizam a educação do cônjuge como instrumento, seguindo os pressupostos de Trostel et. al. (2002), os quais afirmam que a escolaridade de um é positivamente correlacionada com a do outro. Porém, eles mostram que ao utilizar esta variável o trabalho encontra duas limitações: primeiro, o tamanho da amostra é reduzido consideravelmente, dada a restrição de serem incluídos apenas indivíduos casados com informação disponível; segundo, a decisão de casar-se pode não ser aleatória, ou seja, é possível que algumas informações não observáveis que afetam a escolha de casar também possa ter impacto sobre os rendimentos individuais.

Além do acirrado debate de cunho econométrico, há abordagens que usam reformas educacionais como instrumento para tratar a questão de endogeneidade, focando aspectos específicos da distribuição educacional ou de impactos mais amplos de mudança de política sobre a distribuição de renda ou múltiplas *cohorts* de análise. Devereux e Fan (2011), por exemplo, usam IV para investigar os efeitos causais da educação sobre os rendimentos de indivíduos nascidos entre 1970 e 1975, período de larga expansão educacional no Reino Unido. Concluíram que a idade média de educação completada cresceu em um ano de escolaridade no espaço dessas *cohorts*, que resultou em aumentos de salários médios, consequentemente, elevação do retorno da educação.

Muito embora o método de variáveis instrumentais (VI) tenha sido mais frequentemente utilizado para estimar os retornos da educação, tem emergido forte debate na literatura atual acerca da adequabilidade desta abordagem (HECKMAN; LOCHNER; TODD, 2008). Heckman e Urzua (2009) dispõem vários problemas associados com estimação por VI, quais sejam: deficientes instrumentos provêm estimativas viesadas; estimativas de VI advêm de fortes restrições tomadas a priori sobre os dados; em modelos heterogêneos, diferentes instrumentos fornecem estimativas distintas; estimativas de VI, dependendo dos instrumentos utilizados e hipóteses feitas, gerarão diferentes estimativas do retorno da educação, as quais podem levar a incorretas interpretações. Alguns autores como Carneiro e Heckman (2002) e Heckman, Lochner e Todd (2006) relaxam as hipóteses de viés de habilidade e endogeneidade da educação, argumentando que são necessários instrumentos mais confiáveis do que os disponíveis, principalmente quando se utiliza dados a partir de pesquisas censitárias. Em vista disso, outras abordagens têm sido propostas para corrigir tais distorções, por exemplo, através de avaliações de políticas com utilização de efeitos de tratamento em bases econométricas (HECKMAN e VYTLACIL, 2005).

Alguns trabalhos recentes têm relaxado a abordagem minceriana padrão. Henderson, Polachek e Wang (2011) empregam regressão Kernel não-paramétrica para

examinar as diferenças nas taxas de retorno da educação, tanto intra quanto inter grupos, concluindo, ao aplicar sua análise para os Estados Unidos, com resultados contrários aos de muitos outros, tais como, o retorno da educação de indivíduos pretos ser maior do que o de brancos, assim como o dos mais jovens também ser maior.

Diversos autores têm incorporado abordagens de programação dinâmica com escolha discreta para modelar escolaridade e decisões de mercado de trabalho relacionado. Keane e Wolpin (1997) estimaram um modelo dinâmico de investimento em capital humano da escolaridade, decisões de emprego e de profissão, onde o viés de heterogeneidade de habilidade desempenha um papel em todas as três opções. Belzil e Hansen (2002) estimam um modelo de programação dinâmica onde indivíduos diferem no mercado e na habilidade resultante da escolaridade, e relaxam a hipótese de retorna marginal constante da escolaridade. Eles encontram evidências sobre viés de capacidade entre indivíduos e, talvez mais importante, que a relação entre o logaritmo dos rendimentos e escolaridade é altamente não-linear, de forma que, métodos de estimação que não permitem a obtenção de estimativas flexíveis do retorno da escolaridade levará a estimativas inconfiáveis desse retorno

## 4. Resultados

### 4.1 Descrição da Amostra

As médias das variáveis utilizadas para inferência referentes ao Brasil, regiões NE e SSE e estados do Ceará e Rio de Janeiro estão explicitadas na tabela 1. Como se observa, o Brasil apresenta uma renda média (Y) de R\$ 877.62, resultado de um diferencial de renda entre as regiões, visto ser a média de renda dos indivíduos residentes na região Nordeste (NE) aproximadamente 40% inferior à verificada para as regiões Sul-Sudeste (SSE). E ao se comparar os estados representativos, Ceará (CE) e Rio de Janeiro (RJ), o diferencial em favor do segundo é ainda maior. Referentes à variável representativa do capital humano, anos de estudo do indivíduo (E), a diferença entre NE e SSE e entre CE e RJ é de 16% e 10%, respectivamente, resultados estes que deverão ser revertidos nas taxas de retorno. Além disso, dentre as estratificações da amostra, o NE apresenta a maior proporção de indivíduos do sexo masculino, no valor de 61%, porém é a região com a menor quantidade de indivíduos residentes em área urbana, 83%.

Tabela 1 – Médias das Variáveis na Amostra por Região e Estado, 2009

Variáveis	Brasil	NE	SSE	CE	RJ
Y	877.62	623.45	995.65	615.12	1015.56
E	9.19	8.33	9.7	8.72	9.6
I	27.24	27.26	27.21	26.83	28
G	0.58	0.61	0.57	0.59	0.56
R	0.46	0.27	0.60	0.30	0.53
U	0.88	0.83	0.91	0.87	0.96
Total	59073	20085	30709	4515	4096

Fonte: Elaboração própria a partir de dados da PNAD (2009).

## 4.2 Resultados das Estimacões da Equação de Rendimentos

Para atender um dos objetivos propostos por este trabalho em verificar qual modelo, endógeno ou VI, possui o melhor poder preditivo, faz-se necessário, a priori, escolher quais instrumentos serão utilizados no segundo método. Com isso, um vetor de variáveis foi construído para servir de instrumento para a variável anos de estudo, a qual é considerada endógena em uma equação de rendimentos. Tais variáveis são: número de componentes da família e se o indivíduo possui acesso a internet, além das variáveis explicativas. A expectativa é que a primeira exerça uma correlação negativa, por comprometimento do orçamento doméstico, e a segunda positiva por complementaridade à exigência do nível educacional. A tabela 2 mostra as estimativas que fornecem indícios sobre a validade destes instrumentos, uma vez que, independentemente de onde se encontra a amostra, os coeficientes com os sinais esperados são estatisticamente significantes, juntamente com a estatística F, indicando que essas variáveis são correlacionadas com a variável endógena e que podem ser utilizadas no modelo de VI.

Tabela 2 – Validade dos Instrumentos

Variáveis	Brasil	NE	Ceará	Sul-SE	RJ
Nº componentes	-0.3295 (0.000)	-0.2931 (0.000)	-0.3009 (0.000)	-0.3357 (0.000)	-0.3405 (0.000)
Internet	4.2291 (0.000)	4.6244 (0.000)	4.0141 (0.000)	3.7848 (0.000)	3.8836 (0.000)
Const.	7.9834 (0.000)	7.3078 (0.000)	7.9571 (0.000)	8.5156 (0.000)	8.3360 (0.000)
R <sup>2</sup>	0.3429	0.3429	0.2977	0.3149	0.2914
F	16208.91 (0.000)	5701.24 (0.000)	1018.23 (0.000)	7301.71 (0.000)	847.31 (0.000)

Fonte: Elaboração Própria

Nota: p-valores entre parênteses.

Constatada a validade estatística dos instrumentos, estimou-se a equação minceriana com erros robustos para heterocedasticidade (correção de White) para os dois métodos, cujos resultados estão expostos na tabela 3. Pode-se observar que o método de Heckman aplicado ao modelo endógeno possui o menor erro quadrático médio (MSE) independente da estratificação da amostra, ou seja, o método que, a priori, gera estimador viesado possui o melhor ajuste do que o modelo VI. Isso implica que os trabalhos na literatura que procuram corrigir o problema da endogeneidade através da utilização de instrumentos acabam incorrendo em estimativas menos precisas para fins preditivos. Tal ocorrência deve-se à qualidade dos instrumentos que são utilizados, embora se enfatize que estes foram escolhidos levando em consideração as limitações da PNAD para 2009. Estes resultados reforçam a precariedade de variáveis disponibilizadas para uma estimação de um modelo mais sofisticado e alertam para a necessidade de ampliar o conteúdo das amostras.

Tabela 3 – Estimação da Equação de Rendimento através dos Modelos endógeno e VI

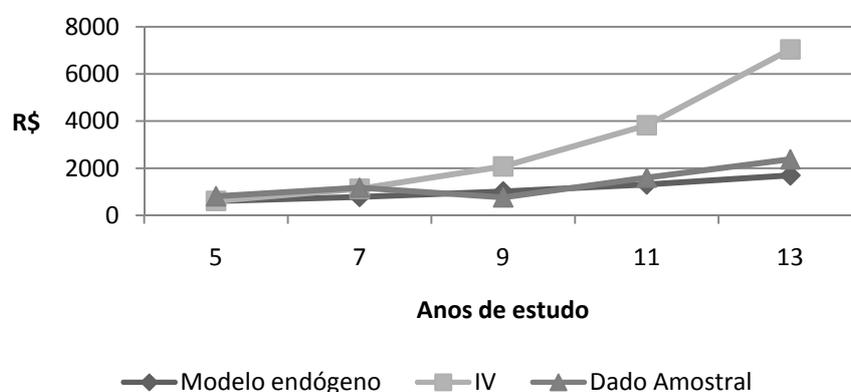
Variáveis	Modelo Endógeno					VI				
	Brasil	NE	Ceará	Sul-SE	RJ	Brasil	NE	Ceará	Sul-SE	RJ
E	0.1285 (0.000)	0.1265 (0.000)	0.1236 (0.000)	0.1239 (0.000)	0.1179 (0.000)	0.3050 (0.000)	0.3476 (0.000)	0.3497 (0.000)	0.2653 (0.000)	0.2170 (0.000)
I	0.1200 (0.000)	0.1474 (0.000)	0.0981 (0.000)	0.1222 (0.000)	0.2135 (0.000)	0.1831 (0.000)	0.2838 (0.000)	0.1463 (0.000)	0.1675 (0.000)	0.4266 (0.000)
$I^2$	-0.001 (0.000)	-0.001 (0.000)	-0.001 (0.024)	-0.001 (0.024)	-0.002 (0.000)	-0.001 (0.000)	-0.003 (0.000)	-0.001 (0.000)	-0.001 (0.000)	-0.005 (0.000)
G	0.5694 (0.000)	0.6050 (0.000)	0.5242 (0.000)	0.5788 (0.000)	0.6660 (0.000)	1.2705 (0.000)	1.6279 (0.000)	1.4896 (0.000)	1.149 (0.000)	1.3295 (0.000)
R	0.2324 (0.000)	0.1365 (0.000)	0.1129 (0.000)	0.1570 (0.000)	0.1267 (0.000)	0.1490 (0.000)	0.0642 (0.000)	0.0436 (0.112)	0.0855 (0.000)	0.0679 (0.000)
U	0.2554 (0.000)	0.3131 (0.000)	0.5091 (0.000)	0.1011 (0.000)	0.0370 (0.369)	0.1075 (0.000)	0.1345 (0.000)	0.3858 (0.000)	-0.021 (0.000)	-0.094 (0.055)
$\varphi$	6.9363 (0.000)	4.6975 (0.000)	3.9223 (0.000)	11.636 (0.000)	138.85 (0.000)	37.780 (0.000)	34.840 (0.000)	36.894 (0.000)	46.168 (0.000)	411.52 (0.000)
Const.	0.2495 (0.146)	0.2492 (0.398)	1.1121 (0.079)	-0.899 (0.000)	-39.18 (0.000)	-12.58 (0.000)	-14.43 (0.000)	-12.93 (0.000)	-13.94 (0.000)	-123.4 (0.000)
R <sup>2</sup>	0.3580	0.3250	0.3174	0.3503	0.3380	0.2054	0.1308	0.1373	0.2207	0.2442
MSE	0.6541	0.7172	0.7234	0.5655	0.5560	0.7285	0.8156	0.8161	0.6197	0.5936
F	3683.97 (0.000)	1123.9 (0.000)	231.62 (0.000)	1835.3 (0.000)	233.50 (0.000)	17380 (0.000)	5407.5 (0.000)	1160.3 (0.000)	8692.9 (0.000)	1163.7 (0.000)

Fonte: Elaboração própria

Nota: p-valores entre parênteses.

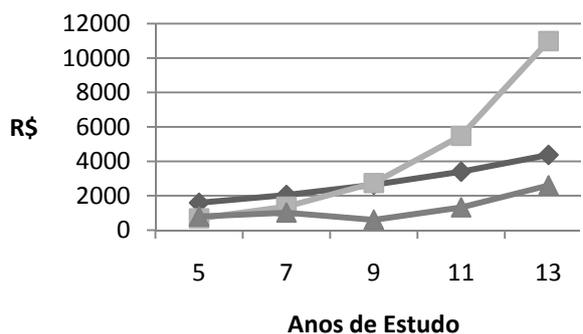
Além disso, essa tabela mostra que o coeficiente que representa a razão inversa de Mills é estatisticamente significativa, indicando que a não inclusão desta variável em uma equação de rendimentos fornecerá estimativas viesadas. Outro ponto que pode ser destacado é que quando se utiliza a estimação pelo método de VI a maioria dos coeficientes das variáveis aumentam em magnitude, inclusive o retorno da educação, o qual apresentou um aumento de, no mínimo, 100%, superestimando assim, as previsões dos retornos sobre os rendimentos. A comparação para fins preditivos entre esses dois modelos com as observações da amostra é ilustrada nos gráficos abaixo, os quais foram elaborados a partir de um indivíduo representativo que apresenta as seguintes características: homem da raça branca com idade de 35 anos e residente em área urbana; as demais variáveis contínuas foram estipuladas no valor da média. Verifica-se em todos os gráficos que a linha de tendência contento os dados amostrais aproxima-se da linha que representa o modelo endógeno, indicando que este possui um melhor ajuste.

Gráfico 1 – Comparação entre os Modelos - Brasil



Fonte: Elaboração Própria

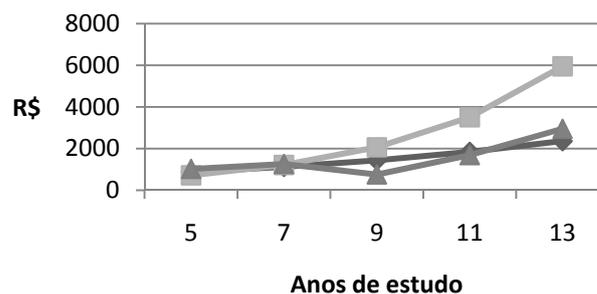
Gráfico 2 - Comparação entre os Modelos – NE



◆ Modelo Endógeno ■ IV ▲ Dado Amostra

Fonte: Elaboração Própria

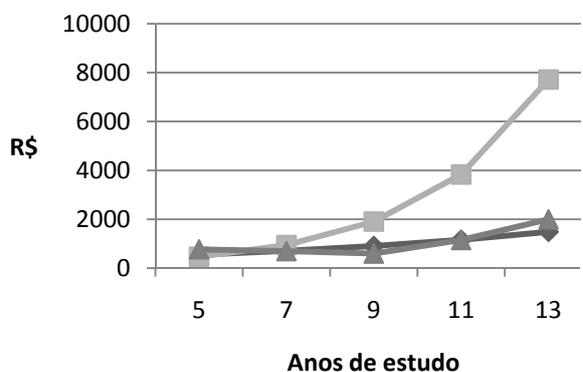
Gráfico 3 - Comparação entre os Modelos – Sul-SE



◆ Modelo Endógeno ■ IV ▲ Dado Amostra

Fonte: Elaboração Própria

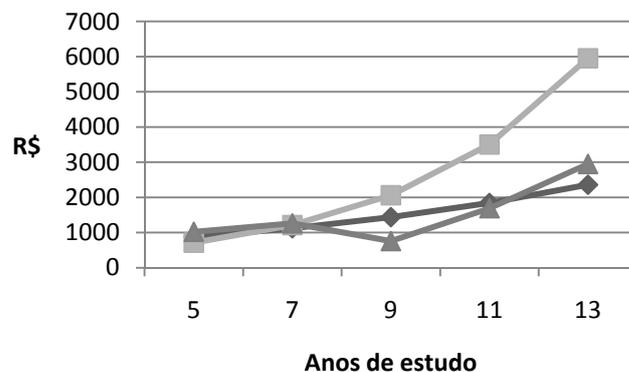
Gráfico 4 - Comparação entre os Modelos – CE



◆ Modelo Endógeno ■ IV ▲ Dado Amostra

Fonte: Elaboração Própria

Gráfico 5 - Comparação entre os Modelos – RJ



◆ Modelo Endógeno ■ IV ▲ Dado Amostra

Fonte: Elaboração Própria

Tendo verificado que o modelo endógeno apresenta um melhor poder preditivo, optou-se por estimar a equação minceriana através da interação entre capital humano e regionalização, visando assim, reduzir o viés de endogeneidade causado tanto pela escolha educacional como o viés de habilidades dos indivíduos. Os resultados podem ser observados na tabela 4, os quais foram estimados com erros robustos e mostram que, não obstante o capital humano (E) ter atendido a expectativa teórica de afetar positivamente os rendimentos, observa-se que os retornos da educação variam com as regiões. Indivíduos residentes na região Nordeste reduzem a magnitude do retorno da educação, em aproximadamente, 12%, enquanto que para as regiões Sul, Sudeste e Centro-Oeste aumentam em respectivamente, 10%, 7%, 11%, respectivamente. Ao restringir a análise por estados selecionados destas regiões, observa-se que o Ceará apresenta o menor retorno, pois seu efeito diminui em 14,46%, enquanto que os estados de Santa Catarina, São Paulo e Mato Grosso do Sul aumentam a magnitude em 12%, 8% e 9%, respectivamente. Isso mostra que um ano de escolaridade tem contribuição menor sobre a variação nos rendimentos para indivíduos de regiões menos desenvolvidas.

Estes resultados reforçam aos encontrados por Bezerra e Ramos (2008), os quais afirmam que a distribuição da educação difere entre as regiões, favorecendo aquelas mais desenvolvidas.

Tabela 4 - Estimação da Equação de Rendimentos por Região e Estado

Variáveis	Coefficiente	Variáveis	Coefficiente
E	0.1667 (0.000)	E	0.1708 (0.000)
$E * D_1$	-0.0199 (0.000)	$E * D_5$	-0.0247 (0.000)
$E * D_2$	0.0176 (0.000)	$E * D_6$	0.0206 (0.000)
$E * D_3$	0.0124 (0.000)	$E * D_7$	0.0143 (0.000)
$E * D_4$	0.0186 (0.000)	$E * D_8$	0.0144 (0.000)
Exp	0.0419 (0.000)	Exp	0.0400 (0.000)
$EXP^2$	0.0001 (0.000)	$EXP^2$	0.0002 (0.000)
G	0.5506 (0.000)	G	0.5542 (0.000)
R	0.1472 (0.000)	R	0.1881 (0.000)
U	0.2018 (0.000)	U	0.2176 (0.000)
$\varphi$	5.6120 (0.000)	$\varphi$	5.7474 (0.000)
Const.	2.0117 (0.000)	Const.	1.9353 (0.000)
R <sup>2</sup>	0.3812	R <sup>2</sup>	0.3615
F	2922.88 (0.000)	F	2508.80 (0.000)

Fonte: Elaboração Própria

Nota: p-valores entre parênteses.

Referente às demais variáveis, tem-se que os coeficientes das variáveis  $exp$  e  $exp^2$  foram significantes, mas a segunda não apresentou sinal esperado<sup>1</sup>, negativo, para indicar o formato parabólico sobre os rendimentos. De acordo com a literatura, existe um diferencial de renda se comparados indivíduos do sexo masculino e feminino e entre indivíduos considerados brancos e os não-brancos. Os resultados aqui encontrados reforçam essa afirmação, pois as variáveis gênero e raça apresentaram sinais positivos, indicando que os indivíduos homens e brancos causam variações de 55% e 15% na renda, respectivamente. Já a variável que representa áreas urbanas observa-se um diferencial de renda entre indivíduos residentes em áreas urbanas e aqueles que residem em áreas rurais, de aproximadamente 20%.

Segundo o modelo teórico de Mincer (1974) o retorno futuro esperado independe do nível de escolaridade escolhido pelo indivíduo. Então para testar essa hipótese optou-se por estimar a equação de rendimentos definidas por faixa de escolaridade, faixas I, II e III, as quais representam os indivíduos que possuem cinco a nove anos de estudo, dez a doze e treze ou mais anos de estudo, respectivamente, cujos resultados constam na tabela 5.

<sup>1</sup> Isso é o resultado de ter-se considerado apenas indivíduos com idade máxima de 35 anos.

Tabela 5 – Estimação da Equação de rendimentos por Faixa de Escolaridade

	Variáveis	Brasil	NE	Ceará	Sul-SE	RJ
Faixa I	E	0.1748 (0.000)	0.1960 (0.000)	0.0908 (0.096)	0.1605 (0.000)	0.1199 (0.000)
	<i>E * EXP</i>	-0.004 (0.000)	-0.005 (0.000)	-0.002 (0.379)	-0.004 (0.000)	-0.003 (0.299)
	EXP	0.0992 (0.000)	0.1146 (0.000)	0.0274 (0.496)	0.1112 (0.000)	0.0727 (0.000)
	<i>EXP</i> <sup>2</sup>	-0.001 (0.000)	-0.001 (0.000)	0.0003 (0.636)	-0.001 (0.000)	-0.001 (0.059)
	G	0.5441 (0.000)	0.6410 (0.000)	0.3021 (0.056)	0.5668 (0.000)	0.4188 (0.000)
	R	0.2209 (0.000)	0.0953 (0.000)	0.1184 (0.000)	0.1361 (0.000)	0.1108 (0.000)
	U	0.2546 (0.000)	0.3662 (0.000)	0.6136 (0.000)	0.0744 (0.000)	-0.015 (0.789)
	$\varphi$	1.3371 (0.150)	1.5452 (0.233)	-5.498 (0.124)	5.6642 (0.000)	-13.48 (0.190)
	Const.	3.2099 (0.000)	2.4888 (0.000)	6.0165 (0.000)	2.2313 (0.000)	8.6610 (0.000)
	F	480.40	178.15	41.02	231.56	26.15
Faixa II	E	0.2515 (0.000)	0.2748 (0.000)	0.1946 (0.086)	0.3335 (0.000)	-0.101 (0.481)
	<i>E * EXP</i>	0.0039 (0.100)	0.0085 (0.106)	0.0144 (0.166)	-0.002 (0.320)	0.0107 (0.157)
	EXP	0.0640 (0.021)	0.0294 (0.637)	-0.075 (0.514)	0.1457 (0.000)	-0.064 (0.433)
	<i>EXP</i> <sup>2</sup>	-0.002 (0.000)	-0.002 (0.000)	-0.002 (0.000)	-0.002 (0.000)	-0.0006 (0.217)
	G	0.6715 (0.000)	0.5527 (0.000)	0.5581 (0.000)	0.6519 (0.000)	0.7226 (0.000)
	R	0.2115 (0.000)	0.0848 (0.000)	0.0342 (0.272)	0.1395 (0.000)	0.0759 (0.000)
	U	0.3011 (0.000)	0.3959 (0.000)	0.5131 (0.000)	0.1696 (0.000)	0.0897 (0.225)
	$\varphi$	40.494 (0.000)	24.929 (0.000)	25.980 (0.147)	46.772 (0.000)	228.11 (0.047)
	Const.	-10.00 (0.000)	-5.990 (0.000)	-5.139 (0.408)	-12.43 (0.000)	-59.55 (0.064)
	F	702.92	169.76	33.32	478.77	52.83
Faixa III	E	0.4116 (0.000)	0.8519 (0.000)	0.3269 (0.124)	0.1649 (0.000)	0.3680 (0.000)
	<i>E * EXP</i>	-0.009 (0.000)	-0.040 (0.000)	-0.009 (0.446)	0.0069 (0.049)	-0.0123 (0.272)
	EXP	0.3244 (0.000)	0.8001 (0.000)	0.2899 (0.120)	0.0528 (0.315)	0.2733 (0.108)
	<i>EXP</i> <sup>2</sup>	-0.005 (0.000)	-0.007 (0.000)	-0.006 (0.000)	-0.004 (0.000)	-0.002 (0.214)
	G	0.5603 (0.000)	0.5077 (0.000)	0.3740 (0.000)	0.5334 (0.000)	0.4339 (0.000)
	R	0.1811 (0.000)	0.2151 (0.000)	0.2264 (0.000)	0.1563 (0.000)	0.0277 (0.672)
	U	0.3651 (0.000)	0.4453 (0.000)	0.8123 (0.000)	0.3329 (0.000)	-0.530 (0.199)
	$\varphi$	101.408 (0.000)	43.470 (0.023)	-3.416 (0.904)	97.038 (0.000)	—
	Const.	-30.06 (0.000)	-19.96 (0.024)	1.5826 (0.888)	-24.90 (0.000)	1.6970 (0.299)
	F	232.13	57.91	23.14	148.76	25.15

Fonte: Elaboração Própria.

Nota: p-valores entre parênteses.

Verifica-se que a taxa de retorno da educação varia de acordo com a faixa de escolaridade, pois na maioria das estratificações o coeficiente desta variável foi estatisticamente significativo e aumenta de acordo com o nível de escolaridade, onde níveis mais elevados de educação fornecem retornos maiores. Porém, ao interagir essa variável com anos de experiência, o coeficiente aproxima-se de zero independente da faixa em que o indivíduo se encontra, confirmando os resultados encontrados por Mincer (1974) que a proporção da renda abdicada por investimentos pós-educação é a mesma em todos os níveis de escolaridade.

Após a comprovação deste resultado, aglutinou-se em uma única equação, seguindo a mesma especificação de Mincer (1974), separando apenas por região e estado e computou-se a mesma variável para observar se os resultados permaneceriam. A tabela 6 mostra que o coeficiente da interação foi significativo, mas também se aproximou de zero em todos os estratos.

Tabela 6 – Estimação da Equação de Rendimentos por Experiência

Variáveis	Brasil	NE	Ceará	Sul-SE	RJ
E	0.2393 (0.000)	0.2699 (0.000)	0.2490 (0.000)	0.2324 (0.000)	0.3195 (0.000)
$E * EXP$	-0.0059 (0.000)	-0.0081 (0.000)	-0.0076 (0.000)	-0.0057 (0.000)	-0.0097 (0.000)
EXP	0.1494 (0.000)	0.1918 (0.000)	0.1751 (0.000)	0.1508 (0.000)	0.2608 (0.000)
$EXP^2$	-0.0023 (0.000)	-0.0033 (0.000)	-0.0031 (0.020)	-0.0023 (0.000)	-0.0041 (0.000)
G	0.5363 (0.000)	0.5682 (0.000)	0.4706 (0.000)	0.5492 (0.000)	0.7105 (0.000)
R	0.2255 (0.000)	0.1229 (0.000)	0.1030 (0.000)	0.1519 (0.000)	0.1029 (0.000)
U	0.2700 (0.000)	0.3279 (0.000)	0.5120 (0.000)	0.1094 (0.000)	0.0497 (0.225)
$\varphi$	4.9743 (0.000)	3.3687 (0.000)	1.9093 (0.140)	9.2930 (0.000)	153.25 (0.000)
Const.	1.2322 (0.000)	1.0356 (0.398)	1.6719 (0.000)	0.2817 (0.201)	-42.878 (0.000)
$R^2$	0.3679	0.3418	0.3324	0.3616	0.3704
F	3348.5 (0.000)	1056.42 (0.000)	214.96 (0.000)	1653.13 (0.000)	222.84 (0.000)

Fonte: Elaboração própria

Nota: p-valores entre parênteses.

Além disso, de acordo com a tabela 5, ao comparar as equações por regiões ou estados dentro de cada faixa de escolaridade, observa-se que os retornos variam dependendo do estrato geográfico e que o estado do Ceará apresentou o menor coeficiente na primeira e na segunda faixa, exceto na terceira faixa, pois este foi insignificante. Isso mostra que indivíduos residentes neste estado e que possuem de cinco a nove anos de estudo apresentam retornos sobre a variação nos rendimentos de aproximadamente 9%, enquanto que se o indivíduo reside no Rio de Janeiro este valor aumenta para 12%.

De acordo com esta tabela, o estado do Ceará apresenta o menor diferencial de renda entre indivíduos do sexo masculino e feminino nas três faixas, no valor de aproximadamente 30%, 55% e 37%, respectivamente. Porém, é o estado com a maior variação nos rendimentos entre indivíduos residentes em áreas urbanas e rurais.

## Conclusões

Este trabalho contribui com o debate existente na literatura econômica sobre o problema da endogeneidade da educação em uma equação de rendimentos individuais ao identificar, através de aplicativos empíricos, o modelo que tem o melhor poder preditivo. Aliado a isso, amplia-se esse debate ao permitir que equações mincerianas provenham retornos variáveis do capital humano através da interação desta variável com a regionalização – grandes regiões e estados selecionados de cada uma destas.

Inicialmente, para consecução desses objetivos, foram testados dois métodos de estimação para a equação minceriana – modelo endógeno e o procedimento de variáveis instrumentais – para se inferir sobre qual deles possui o melhor poder preditivo; o primeiro, decorrente da endogeneidade da variável educação, e o segundo, modelo IV, dependente da qualidade dos instrumentos. Dos resultados, observou-se que o método de Heckman aplicado ao modelo endógeno possui o menor erro quadrático médio (MSE), ou seja, este modelo, sabido a priori ser viesado, se ajusta melhor do que o modelo IV, indicando que, os trabalhos na literatura que procuram corrigir o problema da endogeneidade através da utilização de instrumentos incorrem em estimativas menos confiáveis. Isso acontece devido a restrições que ocorrem quanto às fontes de dados disponíveis.

Após a verificação que o modelo endógeno apresenta um melhor poder preditivo, optou-se por estimar a equação minceriana através da interação entre capital humano e regionalização, visando assim, reduzir o viés de endogeneidade causado tanto pela escolha educacional como o viés de habilidades dos indivíduos. As estimativas permitiram concluir que as taxas de retorno da educação variam de acordo com as regiões brasileiras, sendo a região Nordeste a que apresentou o menor valor dentre elas, correspondente a 14,68%. Comparando estes valores com a região Centro-Oeste, região que apresentou a maior taxa período em análise, o diferencial entre os retornos da educação causada pela diferença regional chega a 26%. Ao restringir a análise por estados selecionados para cada região, observa-se que o retorno para o estado do Ceará está próximo ao encontrado para o NE, no valor de 14.61%, porém ao comparar ao estado de Santa Catarina o diferencial atinge 31%. Isso mostra que indivíduos advindos de regiões com distinto grau de desenvolvimento econômico não acumulam igualmente o mesmo nível de capital humano, portanto, capta-se parte do viés de habilidade.

Além disso, testou-se a hipótese de Mincer de (1974) que o retorno futuro esperado independe do nível de escolaridade escolhido pelo indivíduo e dos resultados se pode verificar através da interação entre anos de estudo e experiência, que o coeficiente desta variável aproxima-se de zero independente da faixa em que o indivíduo se encontra, confirmando os resultados encontrados por Mincer (1974) que a proporção da renda abdicada por investimentos pós-educação é a mesma em todos os níveis de escolaridade.

Portanto, estes resultados reforçam a precariedade de variáveis disponibilizadas para uma estimação de um modelo mais sofisticado e alertam para a necessidade de ampliar o conteúdo das amostras, pois abstraindo-se de problemas de acesso a dados ser um bem público – que muitas agências locais ignoram – agências responsáveis pelo desenho e coleta de dados deveriam compartilhar com a comunidade científica sobre a necessidade de informações a serem geradas em uma *survey* para atender as ações de políticas locais com diagnósticos mais precisos. A exemplo da PNAD que reproduz sistematicamente as mesmas informações sem atender a evolução de mudanças estruturais que ocorrem na economia, e mais rapidamente nas transformações metodológicas das ciências. A inexistência de variáveis instrumentais adequadas é um claro exemplo disso, mas que poderia ser sanado com desenhos

apropriados das *surveys*. Uma mínima análise de cunho intergeracional ou extração de *cohorts* locais que poderiam ser de grande valia na endogeneidade da educação é inviabilizada por indisponibilidade de informação.

O direcionamento de estudos que buscam encontrar a taxa de retorno da educação através de um único número médio foca ações de política com impacto genérico e universal. Para países pobres ou em desenvolvimento com orçamentos restritos não devem apenas esboçar políticas universais, mas sim, concentrar-se em populações locais determinada por suas *cohorts* prioritárias, tais como, classe de indivíduos pobres, alunos do segundo grau evadidos da escola, dentre muitos outros.

Além disso, em termos de política, cabe mencionar que o risco associado com os diferentes níveis de obtenção educacional e o grau de aversão ao risco individual afeta na escolha do nível educacional, pois a presença de risco em retorno da educação gera motivo suficiente para a necessidade de intervenção governamental no setor educacional. Se os indivíduos escolhem níveis educacionais sub-ótimos devido ao elevado grau de aversão ao risco, então, políticas governamentais são necessariamente clamadas.

## Referência Bibliográfica

- ARIAS, O; MCMAHON, W. Dynamic Rates of Return to Education in the U.S. **Economics of Education Review**. v.20, n.1, p. 121-138, 2001.
- ASHENFELTER, O.; KRUEGER, A. Estimates of the Economic Return to Schooling from a New Sample of Twins. **The American Economic Review**. v. 84, n. 5, p. 1157-1173, 1994.
- BARRO; R. P., FRANCO; S., MENDONÇA; R. **A recente queda da desigualdade de renda e o acelerado progresso educacional Brasileiro da última década**. (Texto para discussão n. 1304). Rio de Janeiro: IPEA, 2007.
- BECKER, G. S.; CHISWICK, B.R. Education and the Distribution of Earnings. **American Economic Review**, v. 56, p. 358-369, 1966.
- BEN-PORATH, Y. The Production of Human Capital and the Life Cycle of Earnings. **Journal of Political Economy**, v. 75, n.4, p. 352-365, 1967.
- BELZIL, C. The return to schooling in structural dynamic models: a survey. **European Economic Review**, v.51, p.1059-1105, 2007.
- BELZIL C; HANSEN J. Unobserved Ability and the Return to Schooling. **Econometrica**, v.70, p.2075-2091, 2002.
- BEZERRA; F. B., RAMOS; F.S. **Acesso à educação: houve redução das disparidades regionais e estaduais? Brasil e Nordeste 1981-2002**. Pernambuco: UFPE, 2008
- BOUDARBAT. B; LEMIEUX, T.; RIDELL, W. C. **The Evolution of the returns to Human Capital in Canada, 1980-2006**. (Working Paper n.1) University of British Columbia, 2008.
- BUDRIA, S.; PEREIRA, P.T. **Educational Qualifications and Wage Inequality: Evidence for Europa**. (Discussion Paper, 1763) IZA. Set. 2005.
- CAMERON, A. Colin; TRIVEDI, Pravia K. **Microeconometrics: Methods and Applications**. Cambridge University Press, Cambridge, 2005, 1058p.
- CARD, D. Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. **Econometrica**, v.69, p.1127-1160, 2001.
- CARNEIRO; P., HECKMAN; J. J. The evidence on credit constraints in post-secondary schooling. **Economic Journal**. v. 112 n. 482, p. 705-734. 2002
- DEVEREUX, P.; FAN, W. Earnings Returns to the British Education Expansion. **Economics of Education Review**, (article in press, forthcoming), 15 p., 2011.
- DI PIETRO,G; PEDACE. L. Changes in the returns to education in Argentina. **Journal of Applied Economics**. v. 11, n. 2 , 2008
- FLEISCHHAUER, K. **A Review of Human Capital Theory: Microeconomics**. University of St. Gallen, Discussion Paper n. 01, 2007

HECKMAN, J.J. Sample Selection Bias as a Specification Error. **Econometrica**, v. 47, n. 1, p. 153-161, Jan. 1979.

HECKMAN, J.J. et al. **Earning functions, rates of return and treatment effects: the Mincer Equation and Beyond**. Chicago: USA, Discussion Paper n.1700, 200 p. August, 2005.

HECKMAN; J. J. LOCHNER; L. J., TODD; P. E. **Earnings functions and rates of return**. (Discussion Paper Series). UCD, 2008.

\_\_\_\_\_. Earnings equations and rates of return: The Mincer equation and beyond. In E. A. Hanushek and F. Welch (Eds.), **Handbook of the Economics of Education**. Amsterdam: Elsevier, cap. 7, p. 307- 458. 2006

HECKMAN, J.J; POLACHEK, S. Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship. **Journal of the American Statistical Association** v.69, p. 350-354, 1974.

HECKMAN, J.J.; URZUA, S. **Comparing IV with Structural Models**: What Simple IV Can and Cannot Identify. NBER Working Paper, n. 14.706, 33 p., 2009.

HECKMAN, J.J.; VYTLACIL. Structural Effects, Treatment Effects and Econometric Policy Evaluation. **Econometrica**, v.73, n.3, p.669-738, 2005.

HENDERSON, D.J; POLACHEK, S.W; WANG, L. Heterogeneity in Schooling Rates of Return. **Economics of Education Review** (article in press, forthcoming), 13 p., 2011.

KEANE, M.P; WOLPIN, K.I. The Career Decisions of Young Men. **Journal of Political Economy**, v.105, p.473-522, 1997.

LANGONI, C. G. **Distribuição da renda e desenvolvimento econômico do Brasil**. Rio de Janeiro: Expressão e Cultura, 1973.

LAM, D; LEVINSON, D. Declining inequality of schooling in Brazil and its effects on inequality of wages. **Journal of Development Economics**, n.37, p.199-225, 1992.

LOPEZ, R ; THOMAS, V ; WANG, Y. **Addressing the education puzzle**: The distribution of education and economic reform. World Bank, Policy research working paper, n. 2031, 1998

LOUREIRO, P.; GALRÃO, F. Discriminação no mercado de trabalho: uma análise dos setores rural e urbano no Brasil. **Economia Aplicada**, v. 5, n. 3, p. 519-545, 2001.

MENEZES-FILHO, N. A. **A evolução da educação no Brasil e seu impacto no mercado de trabalho**. São Paulo: USP/ Departamento de Economia, março, 2001.

MINCER, J. Investment in Human Capital and Personal Income Distribution. **Journal of Political Economy**, v.6, n.4, p 281-302, 1958.

MINCER, J. B. **Schooling, experience and earnings**. New York: NBER, 1974. 152p.

PSACHAROPOULOS, G.; PATRINOS, H. A. **Returns to Investment in Education: a Further Update**. World Bank Policy Research Working Paper 2.881, 2002.

RAMOS; L. A desigualdade de rendimentos do trabalho no período pós-real: o papel da escolaridade e do desemprego. **Economia Aplicada**. São Paulo: v. 11, n. 2, p. 281-301, abril-junho 2007

RESENDE, M.; WYLLIE, R. **Retorno para Educação no Brasil**: Evidências Empíricas Adicionais. Texto para discussão n.03, UFRJ, Rio de Janeiro, 2005.

ROUSE, C. Further Estimates of the Economic Return to Schooling from a New Sample of Twins. **Economics of Education Review**. v.18, n.2, p. 149-157,1999.

SACHSIDA; A., LOUREIRO; P. R. A, MENDONÇA; R. J. C. Um Estudo Sobre Retorno em Escolaridade no Brasil. **RBE**. Rio de Janeiro: Abril/Jun. 2004

SALVATO, M. A.; SILVA, D. G. **O Impacto da educação nos rendimentos do trabalhador**: uma análise para a região metropolitana de Belo Horizonte. 2008. Disponível em: < [http://www.marciosalvato.com/pdf/marcio\\_denis.pdf](http://www.marciosalvato.com/pdf/marcio_denis.pdf)> Acesso em: 10 Jul. 2011.

SAPELLI, C. **Los Retornos a la Educación en Chile**: Estimaciones por Corte Transversal y por Cohortes. Documento de Trabajo, n. 349. ISSN: 0717-7593. 2009

SOARES, R. R.; GONZAGA, G. Determinação de salários no Brasil: Dualidade ou não - linearidade no retorno da educação . **Revista de Econometria**, v.19, n.2, 1999.

SULIANO, D. C.; SIQUEIRA, M. L.. **Um estudo do retorno da educação na região nordeste**: análise dos estados da Bahia, Ceará e Pernambuco a partir da recente queda da desigualdade. (texto para discussão, n. 72) Fortaleza, 2010.

TROSTEL, P, WALKER, I.; WOOLLEY, P.. Estimates of the economic return to schooling for 28 countries. **Labour Economics**, v. 9: p. 1-16. 2002

UEDA, E. M.; HOFFMAN, R. Estimando o retorno em educação no Brasil. **Economia Aplicada**, v.6, n.2, 2002.

WILLIS, R.J; ROSEN, S. Education and Self-Selection. **Journal of Political Economy**, v. 87, p.7-36, 1979.

WOOLDRIDGE, J M. **Introdução à Econometria**: uma abordagem moderna. São Paulo Thomson, 2006.

ZEPEDA et al. **Changes in Earnings in Brazil, Chile And Mexico**: Disentangling The Forces Behind Pro-Poor Change In Labour Markets. (Working Paper, n.51). International Policy Centre for Inclusive Growth (IPC - IG). Mar. 2009