

## RETORNOS DA EDUCAÇÃO NO BRASIL EM ÂMBITO REGIONAL CONSIDERANDO UM AMBIENTE DE MENOR DESIGUALDADE

DANIEL CIRILO SULIANO\*  
MARCELO LETTIERI SIQUEIRA†

### Resumo

Este trabalho estima a taxa de retorno da educação para as duas regiões brasileiras que apresentam fortes disparidades sócio-econômicas: Nordeste e Sudeste. O período compreende os anos de 2001-2006, caracterizado na literatura por uma forte queda da desigualdade de renda no Brasil de acordo com as pesquisas domiciliares. Formas alternativas de estimação aos de Mínimos Quadrados Ordinários são apresentadas. As evidências são de que o retorno da escolaridade ainda se mantém em patamares elevados, não obstante as fortes disparidades sócio-econômicas das áreas geográficas e as diversas formas metodológicas de estimação.

**Palavras-chave:** Educação; Retornos; Queda Recente da Desigualdade; Nordeste; Sudeste.

### Abstract

This paper estimates the rate of return to education for two Brazilian regions that have strong socio-economic disparities: Northeast and Southeast. The period covers the years 2001-2006, characterized in literature by a sharp drop in income inequality in Brazil according to household surveys. Alternative forms of estimation of the OLS are presented. The evidence is that the return to schooling is still at high levels, despite the strong socio-economic disparities in the geographical areas and methodological ways of estimation.

**Keywords:** Education; Returns; Northeast; Southeast; Recent Fall of Inequality.

**JEL classification:** J20, J21, J31

---

\* Analista de Políticas Públicas do Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE). Email: daniel.suliano@ipece.ce.gov.br.

† Professor do Curso de Pós-Graduação em Economia da Universidade Federal do Ceará (CAEN/UFC). Email: marlettieri@gmail.com

## 1 Introdução

Diversos foram os estudos que procuraram evidenciar a importância da educação na explicação dos diferenciais de renda do Brasil (ver, por exemplo, Langoni 2005, Reis & Barros 1990, 1991, Leal & Werlang 1991, 2000, Lam & Levinson 1992, Lam & Shoeni 1993, Menezes Filho et al. 2000, Menezes-Filho et al. 2006, Menezes Filho 2001, Sotomayor 2004)

Ressalte-se ainda que desde a estabilidade alcançada após o plano real, têm-se observado um declínio quase contínuo da desigualdade no Brasil (Ramos 2007). Neste contexto, pode-se dividir esse período em dois: (i) de 1995 a 1999, apesar do controle inflacionário, período caracterizado por forte instabilidade macroeconômica em virtude de sucessivas crises externas; e (ii) de 2001 a 2005, no qual a desigualdade vem declinando de forma acentuada e contínua, atingindo em 2005 o nível mais baixo dos últimos 30 anos (Soares 2006a, IPEA 2006, Franco et al. 2006, Barros, Carvalho, Franco & Mendonça 2007, Veloso & F. 2006). Neste mesmo período, Barros, Franco & Mendonça (2007) mostram que, aliada a queda da desigualdade, medida pelo coeficiente de Gini, ocorreu uma rápida expansão educacional no Brasil.

Por tudo isso, o papel do prêmio à escolaridade, isto é, o adicional salarial que o indivíduo recebe resultante de um ano a mais de estudo, ausulta-se como grande objeto de estudo. No Brasil, inúmeros estudos já estimaram a taxa de retorno da educação através da equação de salários *a lá* Mincer (1974) usando-se diferentes bases de dados e métodos de estimação, além de períodos distintos.

Behrman & Birdsall (1983) é o primeiro estudo que usa uma equação de salários para calcular a taxa de retorno educacional brasileira usando o Censo do IBGE de 1970. Leal & Werlang (1991) também estimam uma equação de Mincer empregando *dummies* para diversos grupos de estudo com dados da Pnad de 1976-1989 e do Censo de 1980. Kassouf (1994, 1998) e Silva & Kassouf (2000) empregam o procedimento de Heckman em dois estágios como forma de corrigir um possível viés de seleção amostral. Os dois primeiros trabalhos usam dados da Pesquisa Nacional de Saúde e Nutrição do IBGE, IPEA e INAN de 1989, enquanto que o último usa dados da Pnad de 1995. Ueda & Hoffmann (2002) usando informações suplementares da Pnad de 1996 dos pais dos indivíduos empregam o método de estimação por variável instrumental, estimadores intrafamiliares, mínimos quadrados (MQO) e um modelo por nível de escolaridade. Sachsida et al. (2004) através de dados empilhados das Pnads de 1992 a 1999 utilizam a metodologia do viés de seleção amostral através dos modelos de Heckman em dois estágios e Garen (1984), além da técnica de pseudo-painel. Por fim, Resende & Wyllie (2006) usando dados da Pesquisa sobre o Padrão de Vida do IBGE de 1996 e 1997 utilizam o procedimento de Heckman em dois estágios incorporando como controle adicional a qualidade da educação.

O presente trabalho buscará complementar a literatura apresentada acima no âmbito regional na medida em que estima a taxa de retorno da escolaridade de duas regiões brasileiras com fortes disparidades sócio-econômicas. Não obstante o uso de diferentes técnicas de estimação, diferentes variáveis explicativas e base de dados diversas distribuídas ao longo dos anos, as evidências são de que a taxa de retorno educacional brasileira ainda é persistentemente elevada. Tais resultados coadunam com as evidências internacionais. Por exemplo, estimativas da taxa de retorno da escolaridade para um conjunto

de países mostram que um ano adicional de estudo no Brasil aumenta, em média, o salário em cerca de 15% colocando o país em nono lugar de um total de 71 (Psacharopoulos & Patrinos 2002).

Assim, um trabalho desta estirpe tem inúmeras justificativas. Como dito, desde a década de 90 houve um progressivo aumento da escolaridade no Brasil. Desta forma, seria importante analisar os retornos educacionais neste contexto. Além disto, quase metade da queda na desigualdade de renda familiar observada de 2001 a 2004 deve-se aos rendimentos do trabalho (IPEA 2006). Outro fator a se ressaltar é que a redução da desigualdade de renda esteve associada tanto a fatores ligados aos rendimentos do trabalho, quanto às transferências governamentais, exercendo estes diferentes impactos nas regiões brasileiras (Hoffman 2006, Silveira Neto & Goncalves 2007).

O fato de distintos fatores exercerem diferentes impactos sobre a distribuição de renda em um período de acelerada queda da desigualdade talvez seja explicada pelo histórico no tocante a existência de desigualdades regionais entre as regiões e os estados brasileiros. De acordo com Barros (2011), as desigualdades regionais no Brasil, sendo o Nordeste mais pobre que o Sudeste, mesmo já existentes, agravou-se mais ainda a partir da chegada da coroa portuguesa ao Brasil no século XIX seguido de outros fatores da época, como a abertura do comércio, a diversificação das atividades produtivas e a aceleração da migração de europeus e asiáticos.

De maneira mais específica, pode-se dizer que a questão regional no Brasil tem início a partir dos trabalhos de GTDN (1967) e Furtado (1985). Desde então, tal problema passou a ser visto como uma busca por uma menor disparidade de renda entre as regiões no qual o processo de industrialização daquelas mais pobres exerceria papel primordial para o tão sonhado desenvolvimento. Assim, de acordo com essa teoria, a saída para superar o atraso relativo do Nordeste e o seu velho modelo agroexportador seria através de estímulos a industrialização da região, setor o qual comandara a dinâmica da região Sudeste à época de seu maior crescimento. Essa tese serviu de base para criação da SUDENE tendo ainda direcionado a maior parte das políticas de desenvolvimento regional no Nordeste a partir da segunda metade do século XX.

Recentemente, alguns autores vêm procurando reformular o conceito de desequilíbrio regional no Brasil. Nesse contexto, a problemática regional brasileira saiu do foco estruturalista e passou a ser analisada sob o enfoque da Teoria Econômica Neoclássica e seus desdobramentos.

É dentro dessa contextualização que Barros (2011) defende que a existência de um problema regional se dá quando as oportunidades para indivíduos de características semelhantes não são as mesmas nas diferentes regiões. Segundo esse argumento, a mera existência de disparidade de renda *per capita* entre regiões não pode ser caracterizada como determinante para uma questão regional.

Além disso, no Brasil, boa parte das desigualdades regionais é explicada a partir da concentração das rendas individuais. Logo, políticas de distribuição de renda teriam mais importância caso o objetivo maior seja reduzir as disparidades regionais (Barros 2011).

Na esteira dessa mesma ótica, Pessoa (2001) caracteriza como problema regional o fato de o diferencial de renda entre trabalhadores com características similares em diferentes regiões explicar uma parcela elevada dos diferenciais de renda *per capita*. Assim, prossegue o argumento, se mesmo após o controle das características dos trabalhadores o diferencial de renda desaparecer, não

existiria um problema regional, mas sim um problema social tal qual a região pobre seria predominantemente povoada por pessoas que detém características de renda baixa.

Além desta introdução, este trabalho contém cinco seções. Na seção 1 é feita uma breve revisão da literatura no que tange a disparidade regional e ao papel da desigualdade de renda brasileira concatenada às taxas de retorno da educação bem como suas implicações para o mercado de trabalho. A seção 2 faz um aparato descritivo da base de dados. A seção 3 é dividida em mais duas sub-seções discutindo-se sobre os diferentes modelos e as diversas variáveis explicativas. Na seção 4 são discutidos os resultados das estimações. Por fim, a seção 5 contém as conclusões do trabalho.

## 2 Breve revisão da literatura

De acordo com Cano (1998), a formação e o estabelecimento de uma divisão inter-regional do trabalho no Brasil acelerou-se a partir da década de 1930 em decorrência do recrudescimento da industrialização e da urbanização neste período. Não obstante as já existentes diferenças históricas entre as regiões, houve a partir daí um processo de exacerbação das disparidades sócio-econômicas, especialmente entre o Nordeste e o Sudeste do país.

Neste contexto, sob forte influência das idéias de Raul Prebisch, Furtado (1985) argumenta que regiões voltadas às atividades primárias exportadoras apresentariam desvantagens em relação aquelas dedicadas a produção de produtos industrializados em razão da deterioração dos termos de troca. No Brasil, o Sudeste foi a região que apresentou as melhores condições estruturais em um particular período de florescimento industrial, enquanto o Nordeste tornou-se mero comprador desses bens ficando, assim, atrofiado em um setor chave no processo de seu desenvolvimento.

Outro ponto importante que prejudicou o desenvolvimento das regiões mais periféricas e alargou mais ainda o fosso entre o centro-sul e o nordeste brasileiro são as diferenças de capital humano (ver, por exemplo, Tabela 3 e Tabela 4 a seguir). De fato, essa parece ser a tônica de Pessoa (2001) na qual argumenta que os diferenciais de renda *per capita* entre regiões existem por características das pessoas que moram na região e não por características da região.

Com efeito, diferenças de capital humano foram geradas na formação das regiões brasileiras sendo provavelmente o componente mais importante na determinação das desigualdades regionais (Barros 2011).

Quanto à questão distributiva dos rendimentos, são os trabalhos de Fishlow (1972) e Langoni (2005) na década de 1970 que iniciam a discussão sob o tema. Não obstante o crescimento da renda real, a publicação do Censo de 1970 revelou um aumento considerável da desigualdade de renda na década de 60 no Brasil (Langoni 2005, Ramos & Reis 2000).

Segundo Langoni, a desigualdade de renda ocorrida entre 1960 e 1970 teria como causa dois fatores. Primeiramente, o papel da interação entre oferta e demanda no mercado de trabalho, na qual a acelerada expansão da atividade econômica conduziu a um aumento da demanda por mão-de-obra qualificada que, aliada a uma oferta relativamente inelástica no curto prazo, teve como consequência uma maior dispersão salarial entre os grupos de trabalhadores com diferentes níveis de qualificação. Assim, a educação seria a variável chave

na explicação do aumento da desigualdade de renda neste período, com ênfase no papel de seus retornos (Langoni 2005).

Fishlow, por sua vez, argumentava que as políticas de estabilização efetuadas no período tiveram papel central no processo de desigualdade, na medida em que a redução do salário mínimo real teria contribuído para a queda do poder aquisitivo de parte dos trabalhadores. Vale também ressaltar que, assim como Langoni, Fishlow enfatizou o papel da educação como fator estruturante no aumento da desigualdade uma vez que os grupos menos afortunados foram os menos privilegiados na distribuição educacional no período 1960-1970 (Fishlow 1972).

Posteriormente, o papel da educação voltou a ser reforçado como fonte de desigualdade salarial durante o decênio 1976-1986. Neste período, a escassez de oferta educacional por parte do setor público, conjugada com a falta de investimentos da iniciativa privada, acabou refletindo-se em elevados retornos ao prêmio de escolaridade no Brasil (Leal & Werlang 1991, 2000).

Outra questão pertinente se dá dentro do comércio exterior em razão da crescente integração econômica entre países e as profundas mudanças tecnológicas, abrindo um amplo leque de discussões sobre as transformações no mercado de trabalho, desde o fim dos anos 80. Países como o Brasil, que abriram tardiamente suas economias, experimentaram efeitos simultâneos do comércio internacional e incorporação de novas tecnologias (Arbache 2001).

Tais eventos estariam provocando mudanças na estrutura de demanda por trabalho e, particularmente, favorecendo trabalhadores qualificados, assim como aumentos dos retornos do capital humano nessas economias. De fato, Arbache (2001) mostra que ocorreu grande elevação dos retornos relativos da educação superior completa e queda dos demais grupos a partir de 1992 no Brasil, período que coincide com as reformas comerciais.

Do ponto de vista da oferta relativa, Andrade & Menezes-Filho (2005) também encontram evidências que a proporção de indivíduos com nível de qualificação intermediária na força de trabalho está crescendo, enquanto a proporção de trabalhadores com baixa qualificação e de indivíduos qualificados vêm-se reduzindo. Assim, apesar do viés de crescimento em relação à demanda por mão-de-obra qualificada, uma maior oferta de mão-de-obra intermediária, associada com uma relativa escassez de pessoas com ensino superior, vem elevando os diferenciais de salários na força de trabalho brasileira.

Por outro lado, outros artigos recentemente publicados na literatura econômica brasileira são enfáticos ao atribuírem papel importante do mercado de trabalho, e em particular a educação dentro desse contexto, na redução contínua e acelerada da desigualdade de renda no Brasil (ver, por exemplo, Soares 2006b, Hoffman 2006, Barros et al. 2006, Barros, Carvalho, Franco & Mendonça 2007, Ferreira et al. 2006).

Neste contexto, apesar do enorme papel das transferências ao longo do período, sua contribuição para a magnitude da redução na desigualdade parece ter tido um papel limitado (Hoffmann 2005, Soares 2006b, Ferreira et al. 2006). Por sua vez, o nível médio de escolaridade da população brasileira tem sido um dos fatores chaves. De fato, o papel da escolaridade vem exercendo forte influência desde 1993 em consequência da persistente redução dos retornos médios à educação (Veloso & F. 2006).

### 3 Análise Descritiva da base de dados

A base de dados utilizada neste trabalho corresponde à Pesquisa Nacional por Amostras de Domicílios (Pnad) do Instituto Brasileiro de Geografia e Estatística (IBGE) dos anos de 2001 a 2006. No caso das estimações, os anos foram empilhados (*pooling*). Para fins descritivos, analisou-se a evolução ano a ano de alguns indicadores relativos à desigualdade, educação e mercado de trabalho. Na tabela a seguir, por exemplo, é apresentado o Índice de Gini do rendimento do trabalho principal a partir de um quadro comparativo entre as regiões Nordeste e Sudeste do Brasil.

**Tabela 1:** Evolução do Índice Gini - Rendimento do Trabalho Principal

Anos	Nordeste	Sudeste
2001	0,60 (0,0038)	0,55 (0,0040)
2002	0,56 (0,0057)	0,52 (0,0042)
2003	0,55 (0,0040)	0,51 (0,0037)
2004	0,56 (0,0057)	0,50 (0,0045)
2005	0,55 (0,0041)	0,50 (0,0036)
2006	0,55 (0,0053)	0,50 (0,0042)

Fonte: Cálculo pelos autores a partir dos dados da Pnad. Obs. Erros-padrão entre parênteses obtidos por *bootstrap*.

Da Tabela 1, pode-se constatar que a região Sudeste começa a série com um menor grau de concentração que a região Nordeste chegando cada uma a reduzir pouco mais de 9% no grau de desigualdade. Ao final da série, o *gap* entre as regiões permanece inalterado o que indica uma melhor situação para o Sudeste quando se considera o Índice de Gini como indicador de desigualdade de renda.

Outra forma de analisar a evolução da desigualdade para as regiões seria através da razão do rendimento do trabalho principal entre o nonagésimo e o décimo percentil da distribuição<sup>1</sup>. De acordo com a Tabela 2, além de o Nordeste apresentar maior dispersão e ter ao final da série um índice superior que no início, no Sudeste ocorre um processo contínuo de menor dispersão.

Em termos educacionais, as Tabela 3 e Tabela 4 apresentam dados referentes ao quantitativo educacional das duas regiões para cada ano da amostra. Pela tabela Tabela 3 observa-se uma pequena evolução das duas regiões em relação à média de anos de estudos dos indivíduos. Mais uma vez, a região Sudeste apresenta vantagem em relação ao Nordeste tendo em conta que sua média ao longo de toda a série chega a ser quase dois anos superior.

A composição educacional, em termos dos distintos níveis de ensino, é outra forma de analisar a oferta educacional de ambas as regiões. De acordo com

<sup>1</sup>Neste caso, foram excluídos os indivíduos de rendimento zero.

**Tabela 2:** Razão entre o Nonagésimo e o Décimo Percentil da Distribuição

Anos	Nordeste	Sudeste
2001	44,2 (0,3664)	25,7 (0,3491)
2002	43,1 (0,5962)	24,6 (0,0000)
2003	43,0 (0,5962)	23,5 (0,0000)
2004	43,5 (0,3681)	21,3 (0,1438)
2005	46,8 (0,3696)	20,9 (0,3506)
2006	45,4 (0,0000)	20,5 (0,1478)

Fonte: Cálculo pelos autores a partir dos dados da Pnad. Obs. Erros-padrão entre parênteses obtidos por *bootstrap*.

**Tabela 3:** Média de Anos de Estudo para Homens entre 24 e 65 Anos de Estudo

Anos	Nordeste	Sudeste
2001	5.2	7.1
2002	5.4	7.4
2003	5.5	7.5
2004	5.7	7.6
2005	5.9	7.8
2006	6.1	8.0

Fonte: Cálculo pelos autores a partir dos dados da Pnad.

**Tabela 4:** Percentual dos Homens entre 24 e 65 anos com Ciclo Escolar Completo

Anos	Nordeste			Sudeste		
	Ensino Fundamental Completo	Ensino Médio Completo	Ensino Superior Completo	Ensino Fundamental Completo	Ensino Médio Completo	Ensino Superior Completo
2001	32.9	21.8	4.7	47.1	30.5	8.8
2002	34.2	23.2	5.2	49.4	33.2	9.7
2003	35.9	24.2	4.7	51.2	34.7	10.0
2004	37.6	25.6	5.3	52.8	36.0	9.6
2005	38.9	27.3	5.2	54.3	38.2	10.2
2006	41.1	28.8	5.5	57.0	40.8	11.1

Fonte: Cálculo pelos autores a partir dos dados da Pnad.

a Tabela 4, é fato que, mesmo que lentamente, existe um progresso educacional nas duas áreas geográficas, principalmente no Sudeste, onde o estoque de homens que completaram algum ciclo escolar avança mais rapidamente. Mesmo assim, esses valores ainda são baixos quando comparados com países desenvolvidos, mesmo se considerássemos os homens que cursaram, sem necessariamente terem completado o ensino superior ou se considerássemos o percentual dos que completaram o ensino médio<sup>2</sup>.

Do lado da demanda, a Tabela 5 apresenta informações no que tange a distribuição ocupacional por ramo de atividade. Como se pode observar, em cada região observa-se um comportamento estanque entre os três ramos de atividades com apenas pequenas oscilações ao longo dos anos.

Em um comparativo inter-regional, observa-se que o setor secundário (comércio e serviços) apresenta uma alocação distributiva similar nas duas regiões, enquanto nos setores primário (agricultura) e terciário (industrial) as diferenças entre elas são significativas. De fato, a agricultura na região Nordeste tem uma participação de 25% no total das ocupações, enquanto que a indústria do Sudeste representa 1/3 da distribuição ocupacional da região.

**Tabela 5:** Distribuição Ocupacional por Ramo de Atividade para Homens entre 24 e 65 anos

Anos	Nordeste			
	Primário	Secundário	Terciário	Outras Atividades /Não declaradas
2001	26,0	49,7	22,5	1,7
2002	26,8	43,2	22,7	7,3
2003	27,1	44,1	21,1	7,7
2004	26,0	43,6	22,3	8,0
2005	26,0	43,1	22,5	8,4
2006	24,2	43,1	23,6	9,0
Anos	Sudeste			
	Primário	Secundário	Terciário	Outras Atividades /Não declaradas
2001	11,2	54,1	32,4	2,3
2002	10,8	45,7	33,2	10,3
2003	11,2	45,9	32,3	10,6
2004	11,1	45,9	33,0	10,1
2005	10,7	45,8	32,9	10,6
2006	10,2	46,2	32,9	10,7

Fonte: Cálculo pelos autores a partir dos dados da Pnad.

## 4 Modelos Econométricos

### 4.1 Modelos

Um grande problema que as equações de salários *a lá* Mincer enfrentam é o viés de variável omitida, como, por exemplo, a habilidade inata. Além disso,

<sup>2</sup>Para mais detalhes desse comparativo internacional e como Brasil está bem aquém da média dos países desenvolvidos, ver Veloso & F. (2006) e Veloso et al. (2009).



como é possível que os indivíduos mais hábeis apresentem custos mais baixos em adquirir educação é também provável que eles escolham estudar mais. Logo, as estimativas de MQO além de medirem os retornos educacionais podem captar ao mesmo tempo os efeitos decorrentes da habilidade das pessoas tornando o estimador viesado e inconsistente (Stock & Watson 2004).

Uma forma de corrigir esse problema é através da estimação de mínimos quadrados em dois estágios (MQ2E) com o uso de variáveis instrumentais (Wooldridge 2002, Stock & Watson 2004). Para que a variável instrumental seja válida, ela deve ser relevante (isto é, correlacionada com o regressor endógeno), além de ser exógena (inexistência de correlação entre o instrumento e o termo de erro). No presente caso, o instrumento a ser utilizado é o número de pessoas na família do indivíduo. De De La Croix & Doepke (2003) constroem um modelo em que a educação e a fertilidade são decisões interdependentes; logo, pais pobres ao decidirem ter mais filhos, levando em conta as restrições orçamentárias, passam a investir menos em educação.

Outra fonte de viés resulta do problema de seletividade amostral. Nesse caso, além de o salário depender da oferta de emprego pode também ser uma função da estratégia de “*job search*” do indivíduo, o que remete ao fato de o mesmo ter implícito um salário de reserva abaixo do qual não aceitaria participar do mercado de trabalho. Dessa forma, o procedimento aqui adotado seria a estimação de uma equação de participação no mercado de trabalho para uma amostra composta por trabalhadores empregados e desempregados, da qual resultaria a geração de uma nova variável *lambda*, conhecida como razão inversa de Mills, passando, então, a ser incluída no modelo original como regressor adicional. Tal metodologia caracteriza o procedimento de dois estágios de Heckman (Heckman 1979).

De outra parte, usando uma extensão da metodologia do viés de seleção amostral, onde não se observa escolhas não-ótimas, pode-se considerar a escolha dos anos de estudo como sendo uma variável contínua e ordenada (Garen 1984). Assim, nesse modelo, controlam-se fatores relacionados aos rendimentos do trabalho por meio da inclusão de uma equação específica para a escolaridade de forma a corrigir problemas de endogeneidade, tendo como subproduto uma variável explicativa *resíduo* e outra interativa entre *resíduo* e *escolaridade*.

Existe ainda um tipo de viés caracterizado pela não observância de atributos específicos dos indivíduos como, por exemplo, esforço, dedicação, persistência, determinação, etc. sendo, portanto, relacionados a componentes idiossincrásicos dos mesmos<sup>3</sup>. Se for esse o caso, deve-se esperar que tais características não se alterem com o tempo e, portanto, tal problema poderia ser contornado a partir do uso da metodologia de dados em painel através do modelo de regressão com efeitos fixos. Como a amostra da Pnad é variável de um ano para outro, não acompanhando os mesmos indivíduos sistematicamente, a alternativa aqui seria o uso da técnica de pseudo-painel, conforme descrito por Deaton (1985).

A vantagem metodológica do pseudo-painel se dá pela não ocorrência de problemas de atritos além da possibilidade de dirimir problemas de erros de medida, já que normalmente empregam-se os dados de forma mais agregada. Neste trabalho, foram utilizados três grupos por agregação, classificados segundo a idade: grupo 24-36 (homens com idade entre 24 e 36 anos); grupo 37-

<sup>3</sup>Habilidade inata é também um atributo idiossincrásico dos indivíduos.

49 (homens com idade entre 37 e 49 anos); grupo 50-65 (homens com idade entre 50 e 65 anos).

Finalmente, com o objetivo de investigar os resultados por ciclo de escolaridade foi estimado um modelo em que a variável contínua *anos de estudo* foi substituída por três *dummies* mensurando os retornos da educação do ciclo completo de ensino, a saber: ensino fundamental (*EF*), ensino médio (*EM*) e ensino superior (*ES*), no qual não ter completado nenhum ciclo escolar foi tomado como categoria de controle. A idéia aqui subjacente é que os ganhos dos agentes que logram êxito em concluir todo um ciclo de estudo são maiores do que aqueles que apenas concluíram algumas séries desse nível (Hungerford & Solon 1987).

#### 4.2 Descrição das variáveis

Com base na amostra total da Pnad foram feitas algumas filtrações de forma a obter uma maior consistência do modelo. Assim, utilizou-se apenas homens com idade entre 24 e 65 anos. A escolha dos homens e não das mulheres se deve a especificidades de ambas as categorias decorrentes, por exemplo, da fertilidade do sexo feminino em função de tarefas reprodutivas assim como discriminação (ver, Leme & Wajman 2001). Por sua vez, a idade mínima de 24 anos é uma praxe de trabalhos no âmbito do mercado de trabalho e que também analisam retornos da educação (ver, por exemplo, Menezes-Filho et al. 2004, Sachsida et al. 2004). A idade máxima de corte de 65 anos segue uma perspectiva alternativa destes trabalhos na medida em que a permanência dos homens no mercado de trabalho é cada vez mais longa em virtude do aumento de sua expectativa de vida o que tem, por consequência, aumentado o tempo laboral.

A inclusão da variável experiência seguiu a abordagem de Heckman et al. (2000), segundo a qual a idade é subtraída dos anos de estudo e dos anos pré-escolares. Além do mais, como nem todos os postos de trabalho são igualmente produtivos, foi incluída a variável tempo de permanência no mesmo emprego como forma de controlar parte da produtividade do trabalhador. Semelhantemente a experiência, Wooldridge (2002) sugere um efeito linear e quadrático da variável.

Em função da forte segmentação e discriminação do mercado de trabalho brasileiro (Barros, Franco & Mendonça 2007), foram acrescentadas *dummies* diferenciando trabalhadores do segmento formal-informal e brancos e não brancos<sup>4</sup>. Similarmente, adicionou-se mais uma *dummy* para diferenciar o setor urbano metropolitano do setor urbano não metropolitano e do setor rural, tendo em conta a falta de convergência de renda dos mesmos dentro da região Nordeste (ver, por exemplo, Hoffman 2006, Barros, Franco & Mendonça 2007). Por sua vez, considerando o papel dos sindicatos na dispersão salarial no Brasil (Menezes-Filho et al. 2002), uma *dummy* '*sind*' é incluída como forma de captar uma diferença de médias entre trabalhadores sindicalizados e não-sindicalizados.

Interações envolvendo as variáveis experiência e raça branca e as variáveis escolaridade, experiência e raça branca (ver Garen 1984), bem como *dummies*

---

<sup>4</sup>Os não brancos são indivíduos que declararam na Pnad serem da raça preta ou parda. Por convenção, os amarelos e os índios foram excluídos da amostra.

referentes à posição de ocupação no trabalho principal<sup>5</sup> e setor de atividade<sup>6</sup> foram também acrescentadas (ver Ueda & Hoffmann 2002). Foi incluída também uma *dummy chefdom* de forma a diferenciar o chefe de domicílio das demais categorias na condição na família. Por fim, com o objetivo de verificar a evolução dos retornos à educação ao longo do tempo, foram acrescentadas cinco *dummies* de ano<sup>7</sup> assim como interações destas com a escolaridade.

A variável dependente, salário real horário, refere-se ao rendimento do trabalho principal, tendo sido deflacionado pelo Índice Nacional de Preços ao Consumidor (INPC) do IBGE. Além disto, de forma a isolar o impacto da variação das horas trabalhadas sobre os rendimentos salariais, dividiu-se o rendimento mensal por quatro, obtendo-se o rendimento semanal do trabalho. Por conseguinte, o rendimento semanal do trabalho foi dividido pelo número de horas trabalhadas por semana, resultando, finalmente, no salário real horário. Assim, o modelo econométrico teria o seguinte formato:

$$\ln w_i = \beta x'_i + \gamma esc_i + \delta Z_i + \varphi residuo_i + \theta residuo \times esc_i + \lambda lambda_i + \varepsilon_i \quad (1)$$

$$x' = (exp, exp^2, perm, perm^2, branca, esc \times exp, exp \times branca, \\ esc \times exp \times branca, ano_i, ano_i \times esc, formal, \sin d, urbmet, \\ urbnaomet, secundario, terciario, contapropria, funcpub, \\ empregador, chefdom)'$$

onde  $w$  é o salário real horário,  $x'$  é um vetor que indica as diversas características observadas dos indivíduos, conforme descrito, e  $esc$  a escolaridade do indivíduo medida em anos de estudo. No modelo por nível de escolaridade, a variável  $esc$  é substituída por três *dummies* para cada ciclo de ensino. Estes atributos individuais e as demais variáveis explicativas que aparecem nos distintos modelos estão enumerados a seguir:

1.  $exp$  – experiência do indivíduo definida como experiência = idade – escolaridade – 6 (medida em anos).
2.  $exp2$  – experiência ao quadrado.
3.  $perm$  – tempo de permanência do indivíduo no mesmo trabalho (medida em anos).
4.  $perm2$  – tempo de permanência do indivíduo no mesmo trabalho ao quadrado.
5.  $branca$  – se o indivíduo se declarou branco ( $branco = 1$  e demais casos = 0).

<sup>5</sup>Foram acrescentadas as *dummies* conta-própria, funcionário público e empregador como forma de captar uma diferença de média entre o grupo de trabalhadores empregado (categoria base). Os trabalhadores na produção para o próprio consumo e trabalhadores na construção para o próprio uso além daqueles que se declararam com rendimentos zero foram excluídos devido a especificidades dessas atividades.

<sup>6</sup>Uma *dummy* para o setor secundário (indústria) e outra para o setor terciário (comércio e serviços) diferencia-se do setor agrícola (grupo de controle).

<sup>7</sup>O ano de 2001 foi tomada como categoria base.

6. *esc×exp* – interação entre as variáveis escolaridade e experiência.
7. *exp×branca* – interação entre as variáveis experiência e raça branca.
8. *esc×exp×branca* – interação entre as variáveis escolaridade, experiência e raça branca.
9. *anoi* – *dummy* para ano *i*, onde *i* = 2002, 2003, 2004, 2005 e 2006 (*anoi* = 1 e 2001 = 0).
10. *anoi×esc* – interação entre o ano *i* e escolaridade.
11. *formal* – se o indivíduo contribuiu para instituto de previdência em qualquer trabalho (*formal* = 1 e trabalhador informal = 0).
12. *sind* – se o trabalhador é sindicalizado (*sind* = 1 e trabalhador não sindicalizado = 0).
13. *urbmet* – indivíduo da região urbana metropolitana (*urbmet* = 1 e rural = 0).
14. *urbnaomet* – indivíduo da região urbana não metropolitana (*urbnaomet* = 1 e rural = 0).
15. *secundario* – trabalhador do setor de comércio e serviços (*secundario* = 1 e primário = 0).
16. *terciario* – trabalhador do setor industrial (*terciario* = 1 e primário = 0).
17. *conta-propria* – se o indivíduo é conta-própria (*conta-propria* = 1 e empregado = 0).
18. *funcpub* – se o indivíduo é funcionário público estatutário ou militar (*funcpub* = 1 e empregado = 0).
19. *empregador* – se o indivíduo é empregador (*empregador* = 1 e empregado = 0).
20. *chefdom* – chefe de domicílio (chefe de domicílio = 1 e demais membros = 0).
21. *residuo* – variável gerada da equação específica para escolaridade no modelo de Garen (1984).
22. *residuo×esc* – interação entre a variável gerada da equação específica para escolaridade no modelo de Garen (1984) e a escolaridade do indivíduo (medida em anos).
23. *lambda* – razão inversa de Mills gerada pelo Modelo de Heckman de dois estágios.
24. *efeitos fixos (Zi)* – conjunto de variáveis não observadas que varia de um grupo para outro no modelo de pseudo-painel<sup>8</sup>.

---

<sup>8</sup>grupo 24-36: homens com idade entre 24 e 36 anos; grupo 37-49: homens com idade entre 37 e 49 anos; grupo 50-65: homens com idade entre 50 e 65 anos.

Quando se considera o modelo por nível de escolaridade, a variável *esc* é substituída por três *dummies* e a equação 1 anterior é substituída pela equação 2 abaixo:

$$\ln w_i = \beta x_i' + \gamma_1 EF + \gamma_2 EM + \gamma_3 ES + \varepsilon_i \quad (2)$$

onde:

1. *EF* – indivíduos com ensino fundamental completo (*EF* = 1 e sem escolaridade = 0, não completaram nenhum ciclo de escolaridade).
2. *EM* – indivíduos com ensino médio completo (*EM* = 1 e sem escolaridade = 0, não completaram nenhum ciclo de escolaridade).
3. *ES* – indivíduos com ensino superior completo (*ES* = 1 e sem escolaridade = 0, não completaram nenhum ciclo de escolaridade).

## 5 Estimacões

As Tabelas 6, 7 e 8 abaixo apresentam, respectivamente, as estimacões dos diferentes modelos para as regies Nordeste, Sudeste e para o Brasil com apenas as regies Nordeste e Sudeste (erros-padrão entre parênteses)<sup>9</sup>.

A princípio, pode-se observar a significância estatística de quase todas as variáveis em todos os métodos de estimacão mostrando, portanto, a relevância da grande maioria das variáveis explicativas bem como a robustez de todos os modelos escolhidos para as distintas estimacões.

É importante também destacar a diferença quando se comparam as duas regies geográficas: o termo de intercepto em todos os modelos da região Nordeste é superior em valor absoluto aos modelos do Sudeste. Dado que este coeficiente representa o salário de um indivíduo na base da estimacão<sup>10</sup>, é fato que esta última região apresenta uma maior convergência salarial entre seus trabalhadores do que a primeira.

Com relação às principais variáveis explicativas<sup>11</sup>, observa-se o seguinte: no caso dos trabalhadores dos setores secundário e terciário, principalmente na região Nordeste, seus diferenciais salariais são bem superiores quando comparados com os trabalhadores do setor primário (agrícola). No caso dos chefes de domicílio, situacão semelhante acontece: quando comparados com suas respectivas categorias de referênci, seus diferenciais salariais são também bem elevados. As *dummies* de posicão do trabalho principal que diferencia conta-própria, funcionário público e empregador dos empregados (categoria

<sup>9</sup>Os resultados das equações de seleçao para o modelo de Heckman bem como as estimacões do primeiro estágio do modelo de variável instrumental e os resultados da equaçao para a escolaridade que gera os resíduos para o modelo de Garen (1984) estão no Apêndice. No modelo de variável instrumental a estatística *F* do primeiro estágio, que fornece uma boa medida da validade do instrumento, mostrou-se altamente significativa para as três áreas geográficas. Os erros-padrão do modelo de Garen (1984) foram obtidos por *bootstrap*.

<sup>10</sup>A categoria de controle nos modelos é dada por um indivíduo não branco e não chefe de domicílio, sem nem um tipo de vínculo empregatício formal e não sindicalizado no ano de 2001. Esse mesmo indivíduo reside na região rural pertencente ao setor agrícola e sendo ainda da categoria empregado. Nos modelos estimados por ciclo escolar a categoria omitida é dada por um homem que frequenta ou frequentou a escola, mas que não completou nenhum ciclo.

<sup>11</sup>O efeito parcial (marginal) exato das principais variáveis explicativas descritas no modelo é obtido através da expressao  $\frac{\partial y}{\partial x} = [\exp(\text{coef}) - 1] \times 100$ , onde *exp* é o número e *e coef* são os coeficientes nas Tabelas 6, 7 e 8.

**Tabela 6:** Estimativas para a Região Nordeste - Pooling (2001-2006)

Variável Dependente: Logaritmo do Salário Real Horário							
Variáveis Explicativas#	MQO	Heckman em 2 Estágios	Pseudo-Painel Efeitos Fixos	Instrumental (IV)	Garen (1984)	Ciclo de Escolaridade	
constante	-1,2312 (0,0247409)	-1,9188 (0,2008244)	-	-1,2772 (0,0789033)	-0,9365 (0,0339277)	-0,6100 (0,0138612)	
esc	0,1367 (0,001998)	0,1621 (0,0101029)	0,1444 (0,0023726)	0,1522 (0,0100911)	0,1071 (0,0043304)	-	
exp	0,0462 (0,0013617)	0,0528 (0,00562)	0,0576 (0,0019935)	0,0454 (0,0036251)	0,0270 (0,0016329)	0,0258 (0,0008478)	
exp2	-0,000512 (0,0000183)	-0,000616 (0,0000768)	-0,000689 (0,0000281)	-0,000442 (0,0000311)	-0,000305 (0,0000222)	-0,000372 (0,0000135)	
branca	0,0318 (0,0121933)	0,0375 (0,0469526)	0,0474 (0,0138149)	0,0684 (0,0107505)	0,0691 (0,0126413)	0,1007 (0,00047403)	
escxexp	-0,0012 (0,0000602)	-0,0014 (0,0002387)	-0,0016 (0,0000802)	-	-0,0001 (0,0000682)	-	
expxbranca	-0,0020 (0,0003664)	-0,0020** (0,0013982)	-0,0021 (0,0004278)	-	-0,0039 (0,0003482)	-	
escxexpxbranca	0,000978 (0,0000382)	0,000950 (0,0001476)	0,000969 (0,0000478)	-	0,000888 (0,0000398)	-	
dummy2002	-0,0024** (0,0111411)	0,4777 (0,1295395)	-	-0,0333 (0,0077887)	0,0019 (0,0125747)	-0,0156* (0,0073504)	
dummy2003	-0,0399 (0,0112292)	0,4553 (0,1333436)	-	-0,1035 (0,008103)	-0,0346 (0,0129849)	-0,0693 (0,0073581)	
dummy2004	-0,0324 (0,0111926)	0,4665 (0,1341654)	-	-0,1006 (0,0086144)	-0,0246 (0,0131971)	-0,0604 (0,0072882)	
dummy2005	0,0021** (0,0111694)	0,5022 (0,1344321)	-	-0,0770 (0,0092684)	0,0148 (0,0137324)	-0,0243 (0,0072588)	
dummy2006	0,0649 (0,0112591)	0,5813 (0,1384928)	-	-0,0184* (0,0103434)	0,0830 (0,0137098)	0,0492 (0,0071519)	
dummy2002xesc	-0,0038 (0,0015788)	-0,0226 (0,007651)	-0,0040 (0,0011661)	-	-0,0044 (0,0018064)	-	
dummy2003xesc	-0,0090 (0,0015812)	-0,0286 (0,0078008)	-0,0134 (0,0011368)	-	-0,0091 (0,0017898)	-	
dummy2004xesc	-0,0085 (0,0015598)	-0,0282 (0,0077546)	-0,0117 (0,0011148)	-	-0,0093 (0,0018477)	-	
dummy2005xesc	-0,0094 (0,0015442)	-0,0292 (0,0077137)	-0,0098 (0,001092)	-	-0,0107 (0,0020482)	-	
dummy2006xesc	-0,0086 (0,0015382)	-0,0291 (0,0078424)	-0,0024 (0,0010737)	-	-0,0108 (0,0019631)	-	

Fonte: Cálculos pelos autores. Número de observações: 128.478. # Os erros padrão são robustos à heteroscedasticidade. Obs. Salvo menção em contrário, todas as variáveis são significativas a 1%. \* Variáveis significativas a 5%. \*\* Variáveis não significativas.

Tabela 6: Estimativas para a Região Nordeste - Pooling (2001-2006) (continuação)

Variável Dependente: Logaritmo do Salário Real Horário	Heckman em Pseudo-Panel		Variável Instrumental (IV)		Garen (1984)	Ciclo de Escolaridade
Variáveis Explicativas#	MQO	Efeitos Fixos	Instrumental (IV)	Garen (1984)	Ciclo de Escolaridade	
perm	0,0154 (0,000579)	0,0159 (0,0006558)	0,0129 (0,0008917)	0,0148 (0,0005493)	0,0153 (0,0006035)	
perm2	-0,000338 (0,0000156)	-0,000342 (0,0000592)	-0,000321 (0,0000185)	-0,000311 (0,0000144)	-0,000311 (0,0000172)	
formal	0,3059 (0,0055561)	0,2947 (0,021301)	0,2247 (0,0188505)	0,3103 (0,0041772)	0,3417 (0,0052488)	
sind	0,0625 (0,005362)	0,0612 (0,0205758)	0,0312 (0,0093574)	0,0571 (0,0060511)	0,0540 (0,0054405)	
urbmet	0,2037 (0,0069137)	0,2287 (0,0278246)	0,0607* (0,0312867)	0,1095 (0,0147229)	0,3047 (0,0067292)	
urbnaomet	0,1226 (0,0062972)	0,1439 (0,0252574)	0,0414* (0,0184563)	0,0215 (0,010901)	0,1904 (0,0064215)	
secundario	0,2017 (0,0058205)	0,2037 (0,0223906)	0,1416 (0,0130411)	0,2225 (0,0052692)	0,2888 (0,0059998)	
terciario	0,2363 (0,0063025)	0,2384 (0,0242117)	0,2361 (0,0066881)	0,2530 (0,0054645)	0,3008 (0,0059151)	
contapropria	-0,0451 (0,0054454)	-0,0447 (0,0208218)	-0,0323 (0,006207)	-0,0347 (0,0050348)	-0,0216 (0,0056221)	
funcpub	0,2749 (0,0096666)	0,2720 (0,0372977)	0,1721 (0,0259598)	0,2036 (0,0093702)	0,2040 (0,0093702)	
empregador	0,6981 (0,0101024)	0,6947 (0,0388786)	0,5814 (0,0295843)	0,6778 (0,0105943)	0,7327 (0,012384)	
chefdom	0,1196 (0,0054248)	0,1549 (0,0232284)	0,1106 (0,0058746)	0,1342 (0,0060224)	0,1322 (0,0052274)	
residuo	-	-	-	-0,0550 (0,0028077)	-	
residuoxesc	-	-	-	0,0076 (0,0001265)	-	
lambda	-	2,9067 (0,7396059)	-	-	-	
grupo24-36	-	-	-	-	-	
grupo37-49	-	-	-	-	-	
grupo50-65	-	-	-	-	-	
EF	-	-	-	-	-	
EM	-	-	-	-	-	0,3219 (0,0066853)
ES	-	-	-	-	-	0,3866 (0,0075055)
						1,0030 (0,0117699)

Fonte: Cálculos pelos autores. Número de observações: 128.478. # Os erros padrão são robustos à heteroscedasticidade. Obs. Salvo menção em contrário, todas as variáveis são significativas a 1%. \* Variáveis significativas a 5%. \*\* Variáveis não significativas.

**Tabela 7:** Estimativas para a Região Sudeste - Pooling (2001-2006)

Variável Dependente:	Logaritmo do Salário Real Horário						
Variáveis Explicativas#	MQO	Heckman em 2 Estágios	Pseudo-Painel Efeitos Fixos	Variável Instrumental (IV)	Garen (1984)	Ciclo de Escolaridade	
constante	-1,0853 (0,0236)	-1,4067 (0,0798)	-	-0,9131 (0,0605)	-0,8079 (0,0276723)	-0,2181 (0,0123)	
esc	0,1521 (0,0018)	0,1629 (0,0049)	0,1613 (0,0020829)	0,1489 (0,0079)	0,1333 (0,0029093)	-	
exp	0,0464 (0,0012)	0,0490 (0,0032)	0,0596 (0,0018102)	0,0288 (0,0021)	0,0249 (0,0013)	0,0212 (0,0007)	
exp2	-0,000496 (0,000016)	-0,000538 (0,000042)	-0,000676 (0,0000253)	-0,000230 (0,000014)	-0,000288 (0,0000178)	-0,000343 (0,000012)	
branca	0,0652 (0,01)	0,0715 (0,025)	0,0282 (0,0108696)	0,1268 (0,0121)	0,0955 (0,0099379)	0,1926 (0,0037)	
escxexp	-0,0020 (0,0001)	-0,0021 (0,0001)	-0,0023 (0,0000752)	-	-0,0005 (0,0000642)	-	
expxbranca	-0,0014 (0,001)	-0,0014** (0,0008)	-0,0012 (0,0003479)	-	-0,0032 (0,0002517)	-	
escxexpxbranca	0,00098 (0,00004)	0,000960 (0,00009)	0,000944 (0,0000432)	-	0,000746 (0,0000352)	-	
dummy2002	-0,0152** (0,0123)	0,2493 (0,053)	-	-0,0368 (0,0067)	-0,0074 (0,0114692)	-0,0180 (0,0064)	
dummy2003	-0,0735 (0,0125)	0,1957 (0,054)	-	-0,1226 (0,0071)	-0,0565 (0,0118627)	-0,0919 (0,0064)	
dummy2004	-0,0422 (0,0125)	0,2295 (0,0544)	-	-0,1243 (0,0075)	-0,0219 (0,012077)	-0,0824 (0,0063)	
dummy2005	0,0023** (0,0124)	0,2667 (0,0532)	-	-0,1007 (0,0081)	0,0264 (0,0105812)	-0,0526 (0,0062)	
dummy2006	0,0567 (0,0125)	0,3275 (0,0543)	-	-0,0616 (0,0092)	0,0938 (0,012469)	-0,0019** (0,0062)	
dummy2002xesc	-0,0020** (0,0015)	-0,0101 (0,0038)	-0,0030 (0,0008179)	-	-0,0025 (0,001561)	-	
dummy2003xesc	-0,0048 (0,0015)	-0,0132 (0,0039)	-0,0118 (0,0008264)	-	-0,0064 (0,0013887)	-	
dummy2004xesc	-0,0084 (0,0015)	-0,0169 (0,0039)	-0,0128 (0,0007913)	-	-0,0103 (0,0014833)	-	
dummy2005xesc	-0,0102 (0,0014)	-0,0183 (0,0038)	-0,0120 (0,0008059)	-	-0,0127 (0,001273)	-	
dummy2006xesc	-0,0110 (0,0014)	-0,0193 (0,0038)	-0,0075 (0,0007989)	-	-0,0151 (0,0016166)	-	

Fonte: Cálculos pelos autores. Número de observações: 135.203. # Os erros padrão são robustos à heteroscedasticidade. Obs. Salvo menção em contrário, todas as variáveis são significativas a 1%. \* Variáveis significativas a 5%. \*\* Variáveis não significativas.



Tabela 7: Estimativas para a Região Sudeste - Pooling (2001-2006) (continuação)

Variável Dependente: Variáveis Explicativas#	Logaritmo do Salário Real Horário				Ciclo de Escolaridade	
	MQO	Heckman em 2 Estágios	Pseudo-Panel Efeitos Fixos	Variável Instrumental (IV)		Garen (1984)
perm	0,0233 (0,0006)	0,0233 (0,0014)	0,0227 (0,0006899)	0,0210 (0,0009)	0,0232 (0,0007222)	0,0240 (0,0006)
perm2	-0,000388 (0,000018)	-0,000387 (0,000044)	-0,000370 (0,0000228)	-0,000402 (0,000022)	-0,000388 (0,0000239)	-0,000391 (0,000021)
formal	0,2434 (0,0045)	0,2435 (0,0111)	0,2410 (0,0051068)	0,2015 (0,0099)	0,2513 (0,0039391)	0,2631 (0,0047)
sind	0,1249 (0,0046)	0,1244 (0,0115)	0,1209 (0,0051225)	0,0903 (0,0088)	0,1210 (0,0036669)	0,1280 (0,0046)
urbmet	0,3918 (0,0073)	0,3780 (0,0186)	0,3414 (0,0081076)	0,2569 (0,026)	0,2477 (0,0119662)	0,4600 (0,0073)
urbnaomet	0,2479 (0,0071)	0,2365 (0,018)	0,2294 (0,0077356)	0,1526 (0,0189)	0,1235 (0,0102633)	0,3072 (0,0071)
secundario	0,0358 (0,0053)	0,0352 (0,0132)	0,0332 (0,0059664)	0,0192 (0,0056)	0,0652 (0,0049965)	0,097 (0,0055)
terciario	0,1171 (0,0054)	0,1169 (0,0136)	0,1069 (0,005944)	0,1278 (0,0069)	0,1426 (0,0050633)	0,1630 (0,0054)
contapropria	0,0786 (0,005)	0,0786 (0,0124)	0,0834 (0,0059684)	0,0523 (0,0078)	0,0843 (0,0062521)	0,0959 (0,0055)
funcpub	0,2008 (0,0081)	0,2006 (0,0201)	0,2220 (0,0086544)	0,1364 (0,0164)	0,1648 (0,00752)	0,1631 (0,0077)
empregador	0,5294 (0,0075)	0,5277 (0,0187)	0,5479 (0,0100991)	0,4493 (0,0197)	0,5139 (0,0082624)	0,5463 (0,0095)
chefdom	0,1631 (0,0047)	0,1836 (0,0125)	0,1603 (0,0050874)	0,1528 (0,0055)	0,1833 (0,0052518)	0,1711 (0,0045)
residuo	-	-	-	-	-0,0762 (0,0025187)	-
residuoxesc	-	-	-	-	0,0075 (0,0001125)	-
lambda	-	1,6625 (0,2726)	-	-	-	-
grupo24-36	-	-	-1,2128	-	-	-
grupo37-49	-	-	-0,0745	-	-	-
grupo50-65	-	-	-0,0147	-	-	-
EF	-	-	-	-	-	-
EM	-	-	-	-	-	0,2533 (0,0051)
ES	-	-	-	-	-	0,3517 (0,0057)
	-	-	-	-	-	0,8165 (0,0078)

Fonte: Cálculos pelos autores. Número de observações: 135.203. # Os erros padrão são robustos à heteroscedasticidade. Obs. Salvo menção em contrário, todas as variáveis são significativas a 1%. \* Variáveis significativas a 5%. \*\* Variáveis não significativas.

**Tabela 8:** Estimativas para o Nordeste e Sudeste - *Pooling* (2001-2006)

Variável Dependente: Logaritmo do Salário Real Horário	MQO	Heckman em 2 Estágios	Pseudo-Painel Efeitos Fixos	Variável Instrumental (IV)	Garen (1984)	Ciclo de Escolaridade
Variáveis Explicativas#						
constante	-1,3053 (0,0177701)	-1,7962 (0,0892971)	-	-1,2292 (0,042592)	-1,0554 (0,0180249)	-0,6062 (0,0090629)
esc	0,1465 (0,0015071)	0,1648 (0,005034)	0,1564 (0,001643)	0,1599 (0,0068695)	0,1281 (0,0024731)	-
exp	0,0453 (0,0009642)	0,0493 (0,0029226)	0,0583 (0,0013474)	0,0340 (0,0013748)	0,0261 (0,0008891)	0,0239 (0,0005499)
exp2	-0,000495 (0,0000131)	-0,000559 (0,0000394)	-0,000681 (0,0000189)	-0,000314 (0,0000118)	-0,000294 (0,0000118)	-0,000361 (0,00000895)
branca	0,0568 (0,0075899)	0,0662 (0,0231877)	0,0476 (0,0083409)	0,1174 (0,0055301)	0,0971 (0,011116)	0,1512 (0,002957)
escxexp	-0,0016 (0,0000464)	-0,0016 (0,0001304)	-0,0019 (0,000055)	-	-0,0003 (0,000054)	-
expbranca	-0,0018 (0,0002451)	-0,0017 (0,0007127)	-0,0017 (0,0002658)	-	-0,0044 (0,000267)	-
escxexpbranca	0,000955 (0,0000292)	0,000922 (0,0000778)	0,000917 (0,0000314)	-	0,000865 (0,0000256)	-
dummy2002	-0,0102 (0,0083637)	0,3577 (0,0589955)	-	-0,0329 (0,005052)	-0,0048 (0,0076554)	-0,0155 (0,0048588)
dummy2003	-0,0576 (0,0085167)	0,3195 (0,0603659)	-	-0,1145 (0,0054)	-0,0475 (0,0071456)	-0,0808 (0,0048807)
dummy2004	-0,0396 (0,0084686)	0,3404 (0,0607268)	-	-0,1102 (0,0053689)	-0,0276 (0,0083955)	-0,0717 (0,0048193)
dummy2005	0,0005 (0,0084916)	0,3765 (0,0601416)	-	-0,0848 (0,0056109)	0,0164 (0,0078964)	-0,0386 (0,0047817)
dummy2006	0,0609 (0,008515)	0,4477 (0,0617011)	-	-0,0352 (0,0059948)	0,0846 (0,008094)	0,0219 (0,0047391)
dummy2002.xesc	-0,0024 (0,0011474)	-0,0167 (0,0037895)	-0,0029 (0,0006744)	-	-0,0029 (0,0010083)	-
dummy2003.xesc	-0,0064 (0,0011529)	-0,0212 (0,003847)	-0,0120 (0,0006714)	-	-0,0073 (0,000943)	-
dummy2004.xesc	-0,0082 (0,0011382)	-0,0231 (0,0038345)	-0,0123 (0,0006511)	-	-0,0094 (0,0010353)	-
dummy2005.xesc	-0,0096 (0,0011294)	-0,0242 (0,0037789)	-0,0109 (0,0006517)	-	-0,0113 (0,0009747)	-
dummy2006.xesc	-0,0101 (0,0011187)	-0,0253 (0,0038233)	-0,0052 (0,0006426)	-	-0,0129 (0,0009409)	-

Fonte: Cálculos pelos autores. Número de observações: 263.681. # Os erros padrão são robustos à heteroscedasticidade. Obs. Salvo menção em contrário, todas as variáveis são significativas a 1%. \* Variáveis significativas a 5%. \*\* Variáveis não significativas.

Tabela 8: Estimativas para o Nordeste e Sudeste - Pooling (2001-2006) (continuação)

Variável Dependente: Variáveis Explicativas#	Logaritmo do Salário Real Horário		Pseudo-Painel Efeitos Fixos	Variável Instrumental (IV)		Garen (1984)	Ciclo de Escolaridade
	MQO	Heckman em 2 Estágios		Instrumental (IV)			
dummysudeste	0,3096 (0,0051722)	0,3115 (0,0139204)	0,1655 (0,0139207)	0,4712 (0,0287863)	0,3859 (0,0058394)	0,3287 (0,0050789)	
escxsudeste	-0,0047 (0,0006919)	-0,0049 (0,0019754)	-0,0063 (0,0007618)	-0,0324 (0,0047002)	-0,0124 (0,0006987)	-0,0002 (0,0006496)	
perm	0,0203 (0,0004302)	0,0203 (0,0012216)	0,0201 (0,0004661)	0,0188 (0,0005107)	0,0197 (0,0004312)	0,0200 (0,0004311)	
perm2	-0,000414 (0,0000132)	-0,000416 (0,0000347)	-0,000399 (0,0000141)	-0,000404 (0,0000137)	-0,000386 (0,0000124)	-0,000385 (0,0000132)	
formal	0,2708 (0,0035016)	0,2709 (0,0107279)	0,2631 (0,0038326)	0,2147 (0,0083786)	0,2775 (0,0028616)	0,3031 (0,0034721)	
sind	0,0883 (0,0035838)	0,0875 (0,0107718)	0,0960 (0,0039)	0,0690 (0,004728)	0,0852 (0,0032193)	0,0864 (0,0035653)	
urbmet	0,2981 (0,004989)	0,3070 (0,0153902)	0,2679 (0,005687)	0,1638 (0,0175348)	0,1595 (0,0099968)	0,3950 (0,0049227)	
urbnaomet	0,1793 (0,0048113)	0,1890 (0,0145177)	0,1702 (0,0052624)	0,0908 (0,0121323)	0,0503 (0,0071495)	0,2545 (0,0047971)	
secundario	0,1242 (0,0041081)	0,1249 (0,0120276)	0,1225 (0,0045091)	0,0760 (0,0065967)	0,1487 (0,0042242)	0,2059 (0,0040947)	
terciario	0,1881 (0,0040353)	0,1890 (0,012641)	0,1802 (0,0044617)	0,1679 (0,0043549)	0,2086 (0,0042563)	0,2499 (0,0040153)	
contapropria	0,0063 (0,0039476)	0,0066 (0,0112187)	0,0155 (0,0042998)	-0,0086 (0,0044625)	0,0153 (0,0038923)	0,0243 (0,0039364)	
funcpub	0,2386 (0,0060291)	0,2372 (0,0191868)	0,2552 (0,0066651)	0,1630 (0,0123422)	0,1872 (0,0058763)	0,1925 (0,0060217)	
empregador	0,5946 (0,0075473)	0,5919 (0,0187762)	0,6111 (0,0079987)	0,5217 (0,0128219)	0,5768 (0,0069614)	0,6204 (0,0076342)	
chefdom	0,1412 (0,0034919)	0,1676 (0,01118)	0,1431 (0,0118881)	0,1341 (0,0038071)	0,1588 (0,004302)	0,1514 (0,0034697)	
residuo	-	-	-	-	-0,0655 (0,0018206)	-	
residuo_xesc	-	-	-	-	0,0075 (0,0000824)	-	
lambda	-	2,1846 (0,3172332)	-	-	-	-	
grupo24-36	-	-	-1,3883	-	-	-	0,2919 (0,0043935)
grupo37-49	-	-	-0,0605	-	-	-	0,3714 (0,0047044)
grupo50-65	-	-	0,0155	-	-	-	0,8834 (0,0068685)
EF	-	-	-	-	-	-	
EM	-	-	-	-	-	-	
ES	-	-	-	-	-	-	

Fonte: Cálculos pelos autores. Número de observações: 263.681. # Os erros padrão são robustos à heteroscedasticidade. Obs. Salvo menção em contrário, todas as variáveis são significativas a 1%. \* Variáveis significativas a 5%. \*\* Variáveis não significativas.

de referência), apresentam também efeitos parciais potencialmente relevantes.

Para distinguir a segmentação de mercado formal-informal optou-se, semelhantemente a Kassouf (1998), diferenciar trabalhadores que pertencem ao setor formal através da contribuição para Instituto de Previdência Social. Assim, tendo como base o questionário da Pnad e usando uma medida ampla de formalidade, trabalhador formal seria aquele que contribuiu para instituto de previdência em *qualquer* trabalho. Os resultados encontrados mostram alta segmentação no mercado de trabalho na medida em que trabalhadores do setor formal, em todos os modelos estimados, ganham, em média, acima de 20% que trabalhadores da informalidade. No tocante a segmentação espacial, não obstante o maior grau de integração entre os meios metropolitano e não-metropolitano e entre as regiões urbana e rural (IPEA 2006, Barros, Franco & Mendonça 2007), observa-se ainda grandes variações intra-regionais entre trabalhadores com características similares.

No caso da variável permanência, todos os modelos mostram que o efeito marginal da variável é bem maior para o Sudeste do que para o Nordeste, apesar do baixo poder explicativo em ambas as regiões. Para a variável experiência, seu poder explicativo acaba se reduzindo levando em conta que a produtividade do trabalhador depende estritamente do seu posto atual de ocupação (IPEA 2006, Barros, Franco & Mendonça 2007). Além disso, a redução da heterogeneidade etária no mercado de trabalho brasileiro talvez explique o baixo poder explicativo dessa variável.

Por fim, e mais importante, segue-se os comentários no que tange à variável *anos de estudo*. Inicialmente, obteve-se o efeito marginal *ceteris paribus* do prêmio à escolaridade com base nos modelos que incluem todos os termos interativos, expresso pela equação abaixo:

$$\frac{\partial(\bullet)}{\partial esc} = \beta_{esc1} + \beta_{esc2} \exp + \beta_{esc3} \exp \times branca + \beta_{esc4} ano_i \quad (3)$$

em que  $\beta_{esc1}$  é o coeficiente da variável *esc*,  $\beta_{esc2}$  o coeficiente do termo de interação *esc*×*exp*,  $\beta_{esc3}$  o coeficiente do termo *esc*×*exp*×*branca* e  $\beta_{esc4}$  o coeficiente do termo de interação *esc*×*anoi*<sup>12</sup>.

Em primeiro lugar, é preciso observar que os coeficientes da variável anos de estudo da região Sudeste são ligeiramente superiores aos da região Nordeste, não obstante eles estejam dentro da mesma margem de confiança o que os torna iguais do ponto de vista estatístico (nos modelos por ciclo de estudo é o Nordeste que apresenta resultados maiores, principalmente para o ensino superior, sendo inclusive diferente do ponto de vista estatístico com relação à região Sudeste).

Nos modelos que incluem conjuntamente as regiões Nordeste e Sudeste (Tabela 8), a *dummy* para a região Sudeste é positiva e altamente significativa. Grosso modo, sem considerar os efeitos ao longo do tempo e assumindo, por hipótese, que as demais variáveis do modelo têm os mesmos retornos entre as regiões, este coeficiente representa a diferença salarial a favor da região Sudeste quando comparado com o Nordeste. Todavia, em termos de retornos de educacionais, de acordo com estas estimativas, um aumento de um

<sup>12</sup>Neste caso, considerou-se apenas homens brancos. Os modelos que incluem conjuntamente as duas regiões apresentam ainda um termo de interação entre a variável anos de estudo e a região Sudeste.

ano a mais de estudo eleva o salário em até 16% na região Nordeste (ver coeficientes dos modelos de Heckman em 2 estágios e variável instrumental). No caso do Sudeste, prevê-se que um aumento de um ano a mais de estudo eleva o salário em até 13% na região. A diferença entre esses dois efeitos, 3%, é o coeficiente do termo de interação *dummy* Sudeste e anos de estudo no método de estimação de variável instrumental (VI).

Além disso, é interessante observar que na estimação por VI os resultados apresentados são condizentes com a literatura teórica e empírica. De fato, como o estimador de MQO super-estima os retornos educacionais em decorrência da auto-seleção dos indivíduos mais hábeis, esperar-se-ia *a priori* um menor valor do coeficiente de VI. É isso que ocorre no modelo estimado para a região Sudeste, mas não no modelo da região Nordeste e nos modelos que incluem conjuntamente as duas regiões.

Não obstante estes dois últimos resultados, outros trabalhos clássicos da literatura empírica apresentam resultados semelhantes de sorte que o estimador de VI acaba sendo superior ao de MQO (ver, por exemplo, Angrist & Krueger 1991). Card (1999) argumenta que estimativas de VI com base nas características das famílias tendem a apresentar valor superior às de MQO.

Por fim, mesmo que marginalmente baixo, convém observar os valores dos coeficientes do termo de interação entre a escolaridade e anos da amostra. Em todos os modelos estimados e para todas as áreas geográficas os coeficientes destes termos iterativos são negativos o que evidencia uma queda nos retornos educacionais no período<sup>13</sup>.

## 6 Conclusões

Este trabalho teve como objetivo estimar a taxa de retorno da escolaridade do Brasil sob a luz de um enfoque regional ao abordar duas regiões com distintos níveis sócio-econômico: Nordeste e Sudeste. O período analisado é caracterizado por uma substancial queda da desigualdade de renda de acordo com os dados das pesquisas domiciliares. Em decorrência de problemas como endogeneidade, seletividade amostral e características não observadas buscou-se aqui alternativas de estimação as de MQO.

Mesmo em um período onde a desigualdade de renda no Brasil esteve entre as mais baixas nos últimos 30 anos aliada ainda a uma rápida expansão educacional, os resultados encontrados, apesar de não comparáveis com outros estudos semelhantes devido às diferentes variáveis explicativas e base de dados, mostram que o prêmio à escolaridade no Brasil ainda se mantém em patamares elevados.

Nesta perspectiva, pelo menos três importantes questões se colocam diante desse problema. Em primeiro lugar, é preciso entender em que medida políticas públicas em termos de educação complementar, através do ensino supletivo, por exemplo, podem suprir a escassez educacional em termos de anos de estudo formal, e, assim, homogeneizar mais a força de trabalho e reduzir a

<sup>13</sup>Nos modelos que incluem conjuntamente as duas regiões (Tabela 8), foram também feitas estimações incluindo um conjunto de interações entre a *dummy* região Sudeste, escolaridade e os anos da amostra de forma a observar a redução do diferencial de retornos entre as regiões. Todavia, além dos valores marginalmente baixos (apenas na terceira casa decimal o valor era diferente de zero), em todos os modelos as estimativas dos coeficientes foram não significativas nem mesmo ao nível de 10%.

taxa de retorno da escolaridade<sup>14</sup>. Além disto, é importante também entender o efeito da composição educacional, e em que medida uma expansão mais acelerada da educação superior pode vir a reduzir o prêmio à escolaridade nos níveis mais elevados de educação.

Por fim, resta saber em que grau a abertura comercial ocorrida no início dos anos 90 está elevando a demanda por mão-de-obra qualificada e provocando, nesta situação, aumentos nos diferenciais salariais.

### **Agradecimentos**

Os autores agradecem as valiosas contribuições de dois pareceristas anônimos da revista sem implicá-los em erros remanescentes.

### **Referências Bibliográficas**

- Andrade, S. A. A. & Menezes-Filho, N. (2005), 'O papel da oferta de trabalho no comportamento dos retornos à educação no Brasil.', *Pesquisa e Planejamento Econômico* 35(2), 189–225.
- Angrist, J. D. & Krueger, A. B. (1991), 'Does compulsory school attendance affect schooling and earnings?', *Quarterly Journal of Economics* 106(4), 979–1014.
- Anuatti-Neto, F. & Fernandes, R. (2000), 'Grau de cobertura e resultados econômicos do ensino supletivo no Brasil', *Revista Brasileira de Economia* 53(4), 165–187.
- Arbache, J. S. (2001), Liberalização comercial e mercado de trabalho no Brasil., *in* 'Microeconomia e Sociedade no Brasil', EPGE-FGV.
- Barros, A. R. (2011), *Desigualdades Regionais no Brasil. Natureza, Causas, Origens e Soluções*, Campus.
- Barros, R. P., Carvalho, M., Franco, S. & Mendonça, R. (2006), 'Uma análise das principais causas da queda recente na desigualdade de renda Brasileira', *Econômica* 8(1), 117–147.
- Barros, R. P., Carvalho, M., Franco, S. & Mendonça, R. (2007), A queda recente da desigualdade de renda no Brasil. r, (texto para discussão, 1.258)., Rio de Janeiro: IPEA.
- Barros, R. P., Franco, S. & Mendonça, R. (2007), A recente queda da desigualdade de renda e o acelerado progresso educacional Brasileiro da Última década, Texto para discussão, 1.304, Rio de Janeiro: IPEA.
- Behrman, J. & Birdsall, N. (1983), 'The quality of schooling: Quantity alone is misleading.', *American Economic Review* 73(5), 928–946.
- Cano, W. (1998), *Raízes da Concentração Industrial em São Paulo*, UNICAMP.

---

<sup>14</sup>Anuatti-Neto & Fernandes (2000) mostram que, de modo geral, o ensino supletivo traz um retorno econômico positivo em termos salariais. Desta forma, a educação complementar seria uma maneira alternativa de resolver gargalos do sistema educacional em ciclos mais inferiores de ensino.

- Card, D. (1999), The causal effect of education on earnings., in 'Handbook of Labor Economics', Vol. 3, Elsevier, pp. 1081–1863.
- De La Croix, D. & Doepke, M. (2003), 'Inequality and growth: Why differential fertility matters.', *American Economic Review* 73(4), 1093–1113.
- Deaton, A. (1985), 'Panel data from time series of cross-sections.', *Journal of Econometrics* 30, 109–126.
- Ferreira, F. H. G., Leite, P. G., Litchfield, J. & Ulyssea, G. A. (2006), 'Ascensão e queda da desigualdade de renda no Brasil.', *Econômica* 8(1), 147–169.
- Fishlow, A. (1972), 'Brazilian size distribution of income.', *American Economic Review* 62(2), 391–402.
- Franco, R. P. B., Carvalho, M. & S. (2006), Pobreza multidimensional no Brasil, Textos para discussão.
- Furtado, C. (1985), *Formação Econômica do Brasil*, Editora Nacional.
- Garen, J. (1984), 'The returns to schooling: A selectivity bias approach with a continuous choice variable', *Econometrica* 52(5), 1199–1218.
- GTDN (1967), 'Uma política de desenvolvimento econômico para o nordeste.', Recife: Sudene.
- Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica* 47, 153–161.
- Heckman, J., Tobias, J. L. & Vytlacil, E. (2000), Simple estimators for treatment parameters in a latent variable framework with an application to estimation the returns to schooling., Technical report, NBER Working Paper 7.950.
- Hoffman, R. (2006), 'Transferências de renda e a redução da desigualdade no Brasil e cinco regiões entre 1997 e 2004', *Econômica* 8, 55–81.
- Hoffmann, R. (2005), 'As transferências não são a causa principal da redução da desigualdade.', *Econômica* 7(2), 335–341.
- Hungerford, T. & Solon, G. (1987), 'Sheepskin effects in the returns to education.', *Review of Economics and Statistics* 69(1), 175–177.
- IPEA (2006), Ipea data, Technical report, Instituto de Pesquisa Econômica Aplicada.
- Kassouf, A. L. (1994), 'The wage rate estimation using heckman procedure', *Revista de Econometria* 14(1), 89–107.
- Kassouf, A. L. (1998), 'Wage gender discrimination and segmentation in the brazilian labor market.', *Brazilian Journal of Applied Economics* 2(2), 243–269.
- Lam, D. & Levinson, D. (1992), 'Declining inequality in schooling in brazil and its effects on inequality in earnings', *Journal of Development Economics* 37(37), 199–225.

- Lam, D. & Shoeni, R. (1993), 'Effects of family background on earnings and returns to schooling: Evidence from Brazil', *Journal of Political Economy* **101**, 710–739.
- Langoni, C. G. (2005), *Distribuição de Renda e Desenvolvimento Econômico no Brasil*, FGV, Rio de Janeiro.
- Leal, C. I. S. & Werlang, S. R. C. (1991), 'Retornos em educação no Brasil: 1976/89', *Pesquisa e Planejamento Econômico* **21**(3), 559–574.
- Leal, C. I. S. & Werlang, S. R. C. (2000), Educação e distribuição de renda, in 'Distribuição de Renda no Brasil', Rio de Janeiro: Paz e Terra.
- Leme, S. C. M. & Wajman, S. (2001), Diferenciais de rendimentos por gênero., in 'Microeconomia e Sociedade no Brasil', EPGE-FGV.
- Menezes Filho, N. A. (2001), Educação e desigualdade., in 'Microeconomia e Sociedade no Brasil', Rio de Janeiro: EPGE-FGV.
- Menezes-Filho, N. A., Fernandes, R. & Picchetti, P. (2006), 'Rising human capital but constant inequality: the education composition effect in Brazil.', *Revista Brasileira de Economia* **60**(4), 407–424.
- Menezes Filho, N. A., Fernandes, R. & Picchetti, P. A. (2000), A evolução da distribuição de salários no Brasil: Fatos estilizados para as décadas de 80 e 90, in 'Desigualdade e Pobreza no Brasil', Rio de Janeiro: IPEA.
- Menezes-Filho, N., Mendes, M. & Almeida, E. S. (2004), 'O diferencial de salários formal-informal no Brasil: Segmentação ou viés de seleção?', *Revista Brasileira de Economia* **58**(2), 235–248,.
- Menezes-Filho, N., Zylberstajn, H. & Chahad, J. P. Z. O. (2002), Os efeitos dos sindicatos sobre o salário médio das firmas Brasileiras., in 'Mercado de Trabalho no Brasil', São Paulo: LTr.
- Mincer, J. (1974), *Schooling, Experience and Earnings*, National Bureau of Economic Research.
- Pessoa, S. (2001), Existe um problema de desigualdade regional no Brasil?, in 'Anais, Encontro Nacional de Economia, 29, Salvador'.
- Psacharopoulos, G. & Patrinos, H. A. (2002), Returns to investment in education: a further update., Technical report, World Bank Policy Research Working Paper 2.881.
- Ramos, L. A. (2007), 'Desigualdade de rendimentos do trabalho no período pós-real: O papel da escolaridade e do desemprego.', *Economia Aplicada* **11**(2), 281–301.
- Ramos, L. & Reis, J. (2000), Distribuição da renda: Aspectos teóricos e o debate no Brasil., in 'Distribuição de Renda no Brasil.', Rio de Janeiro: Paz e Terra.
- Reis, J. & Barros, R. (1990), 'Desigualdade salarial e distribuição de educação: A evolução das diferenças regionais no Brasil', *Pesquisa e Planejamento Econômico* **20**(3), 415–478.



- Reis, J. & Barros, R. (1991), 'Wage inequality and distribution of education.', *Journal of Development Economics* **36**(1), 117–143.
- Resende, M. & Wyllie, M. (2006), 'Retornos para educação no Brasil: Evidências empíricas adicionais.', *Economia Aplicada* **10**(3), 349–365.
- Sachsida, A., Loureiro, P. R. A. & Mendonca, M. J. C. (2004), 'Um estudo sobre o retorno em escolaridade no Brasil.', *Revista de Econometria* **58**(2), 249–265.
- Silva, N. D. V. & Kassouf, A. L. (2000), 'Mercado de trabalho formal e informal: Uma análise da discriminação e da segmentação.', *Nova Economia* **10**(1), 41–78.
- Silveira Neto, R. M. & Goncalves, M. B. C. (2007), Mercado de trabalho, transferências de renda e evolução da desigualdade de renda no nordeste do Brasil entre 1995 e 2005., in 'Anais, Encontro Regional de Economia, 12, Fortaleza'.
- Soares, S. (2006a), A educação no Brasil rural, Technical report, INEP.
- Soares, S. S. D. (2006b), 'Análise do bem-estar e decomposição por fatores da queda na desigualdade entre 1995 e 2004.', *Econômica* **8**(1), 83–115.
- Sotomayor, O. J. (2004), 'Education and changes in brazilian wage inequality, 1976-2001', *Industrial and Labor Relations Review* **58**(1), 94–111.
- Stock, H. J. & Watson, W. M. (2004), *Econometria*, Pearson, Addison Wesley.
- Ueda, E. M. & Hoffmann, R. (2002), 'Estimando o retorno em educação no Brasil', *Economia Aplicada* **6**(2), 209–238.
- Veloso, F., Pessoa, S., Henriques, R. & Giambiagi, F., eds (2009), *Educação Básica no Brasil: construindo o país do futuro*, Elsevier, Rio de Janeiro.
- Veloso, S. F. & F. (2006), 'Intergenerational mobility of wages in Brazil', *Brazilian Review of Econometrics* **26**(2), 181–211.
- Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press: Cambridge.

## Apêndice A RESULTADO DAS EQUAÇÕES PARA O MODELO DE HECKMAN

15

**Tabela A.1:** Equações de Seleção do Modelo de Heckman

Variáveis Explicativas#	Nordeste	Sudeste	Nordeste e Sudeste
constante	0,5710 (0,1087)	0,7899 (0,1369)	0,5992 (0,0841)
esc	0,0500 (0,0082)	0,0538 (0,0095)	0,0513 (0,0061)
exp	0,0338 (0,0062)	0,0328 (0,0075)	0,0319 (0,0048)
exp2	-0,000512 (0,000082)	-0,000506 (0,000098)	-0,000491 (0,000063)
branca	0,0224 (0,0230)	0,0685 (0,0247)	0,0685 (0,0161)
escxexp	-0,0006 (0,0003)	-0,0010 (0,0003)	-0,0006 (0,0002)
urbmet	0,1167 (0,0290)	-0,1932 (0,0457)	0,0579 (0,0231)
urbnaomet	0,0959 (0,0257)	-0,1579 (0,0447)	0,0663 (0,0215)
chefdom	0,1972 (0,0259)	0,2824 (0,0293)	0,2277 (0,0193)
outros rendimentos	-0,0001 (0,0000)	-0,0001 (0,0000)	-0,0001 (0,0000)
dummy2002	1,0702 (0,0343)	1,4352 (0,0543)	1,1976 (0,0283)
dummy2003	1,1661 (0,0377)	1,5962 (0,0680)	1,3040 (0,0320)
dummy2004	1,1936 (0,0382)	1,7079 (0,0783)	1,3435 (0,0330)
dummy2005	1,2001 (0,0377)	1,4216 (0,0520)	1,2850 (0,0302)
dummy2006	1,3438 (0,0439)	1,6330 (0,0680)	1,4465 (0,0364)
Número de observações:	130.519	136.776	267.295

Fonte: Cálculos pelos autores. # Os erros padrão são robustos à heteroscedasticidade. Obs. Salvo menção em contrário, todas as variáveis são significativas a 1%. \*\* Variáveis não significativas.

<sup>15</sup>A notação das variáveis no apêndice a seguir é a mesma que foi descrita ao longo do texto. As novas variáveis que aparecem aqui são: *outros rendimentos* (rendimentos de outras fontes de renda que não o trabalho principal); *nº de membros da família* (número de componentes da família); *mãe viva* (se a mãe do indivíduo é viva). Erros-padrão entre parênteses.

**Apêndice B RESULTADOS DAS ESTIMAÇÕES DO PRIMEIRO  
ESTÁGIO DO MODELO DE VARIÁVEL INSTRUMENTAL**

**Tabela B.1:** Primeiro Estágio dos Modelos de Variável Instrumental

Variáveis Explicativas#	Nordeste	Sudeste	Nordeste e Sudeste
constante	8,1947 (0,0548)	8,0369 (0,0539)	6,5888 (0,0337)
nºde membros da família	-0,1259 (0,0050)	-0,1676 (0,0061)	-0,1352 (0,0039)
exp	-0,3401 (0,0030)	-0,2325 (0,0028)	-0,1738 (0,0019)
exp2	0,002654 (0,000047)	0,000724 (0,000047)	0,000976 (0,000030)
branca	0,9432 (0,0204)	1,4130 (0,0176)	0,6572 (0,0111)
dummy2002	0,2012 (0,0311)	0,1922 (0,0302)	0,1351 (0,0177)
dummy2003	0,2847 (0,0310)	0,3513 (0,0303)	0,2147 (0,0177)
dummy2004	0,4083 (0,0308)	0,4656 (0,0301)	0,2936 (0,0176)
dummy2005	0,5212 (0,0303)	0,6050 (0,0299)	0,3741 (0,0174)
dummy2006	0,6836 (0,0303)	0,7932 (0,0297)	0,4746 (0,0174)
perm	0,0651 (0,0023)	0,0872 (0,0027)	0,0386 (0,0014)
perm2	-0,000633 (0,000056)	-0,000726 (0,000081)	-0,000471 (0,000040)
formal	1,7864 (0,0245)	1,0969 (0,0216)	1,0992 (0,0131)
sind	0,7447 (0,0229)	0,9355 (0,0223)	0,4380 (0,0133)
urbmet	104,8300 (0,0000)	3,1023 (0,0319)	2,3889 (0,0184)
urbnaomet	67,3100 (0,0000)	2,1718 (0,0309)	1,5757 (0,0175)
secundario	1,1183 (0,0262)	0,1228 (0,0269)	0,7294 (0,0155)
terciario	0,1930 (0,0274)	-0,5137 (0,0273)	0,2076 (0,0156)
contapropria	0,5551 (0,0232)	0,6618 (0,0243)	0,2703 (0,0138)
funcpub	2,3989 (0,0430)	1,8290 (0,0379)	1,5697 (0,0258)
empregador	2,6434 (0,0467)	2,1849 (0,0368)	1,5085 (0,0236)
chefdom	0,1516 (0,0233)	0,2727 (0,0221)	0,1134 (0,0134)
dummysudeste	-	-	-4,1638 (0,0148)
escxsudeste	-	-	0,6797 (0,0015)
Teste F	12.417,37	8.358,59	51.603,77
Prob > F	0,0000	0,0000	0,0000
Número de observações:	128.478	135.203	263.681

Fonte: Cálculos pelos autores. # Os erros padrão são robustos à heteroscedasticidade.  
Obs. Salvo menção em contrário, todas as variáveis são significativas a 1%.

**Apêndice C RESULTADO DA EQUAÇÃO PARA ESCOLARIDADE QUE GERA OS RESÍDUOS PARA O MODELO DE GAREN (1984)**

**Tabela C.1:** Resultados da Equação para a Escolaridade que Gera os Resíduos para o Modelo de Garen (1984)

Variáveis Explicativas#	Nordeste	Sudeste	Nordeste e Sudeste
constante	2,9456 (0,0480)	3,6983 (0,0520)	3,1504 (0,0351)
nº de membros da família	-0,2997 (0,0062)	-0,2775 (0,0074)	-0,3125 (0,0048)
branca	1,4205 (0,0254)	1,8858 (0,0213)	1,9176 (0,0158)
urbmet	5,0595 (0,0264)	4,1493 (0,0325)	4,8160 (0,0199)
urbnaomet	3,1537 (0,0257)	3,0711 (0,0320)	3,3763 (0,0195)
chefdom	-0,8341 (0,0291)	-0,7803 (0,0260)	-0,8284 (0,0195)
mãe viva	1,3879 (0,0247)	1,5925 (0,0241)	1,4739 (0,0174)
Wald Chi2()	10.923,11	6.455,91	20.679,88
Prob>Chi2()	0,0000	0,0000	0,0000
R2	0,2605	0,1815	0,2488
Número de observações:	130.519	136.776	267.295

Fonte: Cálculos pelos autores. # Os erros padrão são robustos à heteroscedasticidade. Obs. Salvo menção em contrário, todas as variáveis são significativas a 1%.