

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

A RELAÇÃO ENTRE O DESEMPENHO ESCOLAR E OS SALÁRIOS NO BRASIL

Andréa Zaitune Curi

Orientador: Prof. Dr. Naércio Aquino Menezes Filho

SÃO PAULO
2006

Profa. Dr. Suely Vilela
Reitora da Universidade de São Paulo

Profa. Dra. Maria Tereza Leme Fleury
Diretora da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Joaquim José Martins Guilhoto
Chefe do Departamento de Economia

Prof. Dr. Dante Mendes Aldrighi
Coordenador do Programa de Pós-Graduação em Economia

ANDRÉA ZAITUNE CURI

A RELAÇÃO ENTRE O DESEMPENHO ESCOLAR E OS SALÁRIOS NO BRASIL

Dissertação apresentada ao Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito para a obtenção do título de Mestre em Economia.

Orientador: Prof. Dr. Naércio Aquino Menezes Filho

SÃO PAULO

2006

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Curi, Andréa Zaitune

A relação entre o desempenho escolar e os salários no Brasil / Andréa Zaitune Curi. -- São Paulo, 2006.

71 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2006
Bibliografia.

1. Salários – Brasil 2. Desempenho escolar 3. Educação – Brasil
I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade II. Título.

CDD – 331.220981

A meus pais.

Agradeço ao professor e orientador Naércio Aquino Menezes Filho principalmente, por sua orientação nessa dissertação desde a escolha do tema até a conclusão da mesma. Todos os trabalhos com ele desenvolvidos resultaram em amplo aprendizado.

Agradeço, também, aos professores Reynaldo Fernandes e Eduardo de Carvalho Andrade pelas críticas e sugestões a este trabalho em sua qualificação.

Aos meus colegas de pós-graduação do IPE-USP pelas ajudas nas disciplinas, pela sempre rica troca de idéias, pelo apoio moral fundamental à execução deste trabalho e pela ótima convivência a qual transformou esses anos de mestrado em um período muito agradável. Agradeço a amizade e o companheirismo de pessoas especiais como Ana Carolina Giuberti, Fernanda Cabral dos Santos, Fabio Miessi Sanches, Danilo Ramalho e Ricardo Vieira.

Agradeço ao apoio institucional do IPE-USP e ao apoio financeiro da FIPE e da FAPESP.

Por fim, agradeço a minha família pelo apoio.

RESUMO

O objetivo desse trabalho é analisar a relação entre o desempenho escolar e os salários dos jovens brasileiros. Examinamos se a qualidade do ensino, mensurada pelas notas obtidas por uma geração nos exames de proficiência realizados pelo Instituto Nacional de Estudos e Pesquisas Anísio Teixeira, o INEP, em seu Estado ao término do ensino médio, afeta os salários a serem recebidos por esta geração quando ela estiver no mercado de trabalho, cinco e seis anos depois, respectivamente com dados do Censo Demográfico de 2000 e da Pesquisa Nacional por Amostra de Domicílios, a PNAD, de 2001. A partir de um modelo de pseudo-painel, corrigimos os problemas de viés de seleção gerados pelas migrações e pelo alto nível educacional da amostra selecionada, através do modelo de Roy (1951) aplicado em Dahl (2001). Os determinantes da proficiência escolar, tais como “*background*” familiar, a infraestrutura escolar e o perfil de docentes e diretores também foram analisados. A partir da estimação do modelo em dois estágios, os resultados do primeiro estágio mostram que o desempenho dos alunos em exames de proficiência é positivamente relacionado aos investimentos em educação, como melhor remuneração a professores e diretores, critérios mais rigorosos de seleção desses profissionais, assim como investimentos em infraestrutura, que melhoram a qualidade da escola. Isso aponta para a existência de uma relação entre os recursos destinados à educação e a qualidade da mesma no Brasil. Adicionalmente, os resultados do segundo estágio do modelo mostram que as notas obtidas por uma geração nos testes de proficiência são significantes para explicar os salários futuros da mesma. Dessa forma o estudo confirma a importância de políticas públicas que invistam na qualidade da escola ao invés de políticas destinadas apenas a aumentar os anos de estudo da população.

ABSTRACT

The aim of this paper is to examine the relationship between the school performance and the wages of young Brazilians workers. We examine if school quality, measured by test scores of a generation in SAEB at the end of high school, affect the earnings of this generation when they enter the labor force, five and six years later, with sample of Censo (2000) and PNAD (2001). We use a pseudo-panel model to correct the problems of selection bias, created by migrations and by the high education level of the selected sample, through a Roy model (1951) applied in Dahl (2001). The determinants of school performance, like familiar background, school structure, teacher and director profiles also were analyzed. We conclude that school characteristics are responsible for a good performance of students in tests scores, and that the school performance explains the differences of the earnings of young Brazilians workers.

SUMÁRIO

1	INTRODUÇÃO	12
2	REVISÃO BIBLIOGRÁFICA	13
3	METODOLOGIA ECONOMETRICA	17
3.1	MODELO DE ROY POR DAHL.....	18
3.1.1	Uma extensão do Modelo de Roy: Um Modelo de Mobilidade e Salários	19
3.1.2	Modelo de Viés de Seleção com Múltiplas Escolhas	22
3.1.3	Implementação das Escolhas.....	25
3.1.3.1	Reduzindo o número de funções de correção	26
3.1.3.2	Usando Células de Fluxo Migratório para Possibilidades de Seleção.....	26
3.2	O MODELO ESTIMADO.....	28
4	DADOS.....	29
	Figura 1 – Formação dos Pseudo-Paineis com a PNAD de 2001	31
	Figura 2 – Formação dos Pseudo-Paineis com o Censo de 2000.....	31
	Gráfico 1 – Salário Real Mensal em 2001 por UF.....	33
	Gráfico 2 – Salário Real Mensal em 2000 por UF.....	33
	Gráfico 3 – Salário Real Mensal por UF: PNAD vs Censo.....	34
	Gráfico 4 – Proficiência em Matemática por UF em 1995.....	36
	Gráfico 5 – Proficiência em Português por UF em 1995	36
	Gráfico 6 – Proficiência em Matemática (1995) vs Salário (2001) por UF	38
	Gráfico 7 – Proficiência em Português (1995) vs Salário (2001) por UF	38
	Gráfico 8 – Proficiência em Matemática (1995) vs Salário (2000) por UF	39
	Gráfico 9 – Proficiência em Português (1995) vs Salário (2000) por UF	39
	Gráfico 10 – Proficiência em Matemática (1995) vs Salário (2001) por Célula	40
	Gráfico 11 – Proficiência em Português (1995) vs Salário (2001) por Célula	40
	Gráfico 12 – Proficiência em Matemática (1995) vs Salário (2000) por Célula	41
	Gráfico 13 – Proficiência em Português (1995) vs Salário (2000) por Célula	41
	Tabela 1 – Número de Células com 10 ou mais indivíduos por Pseudo-painel	42

4.1	Análise da Migração no Brasil da Amostra Seleccionada.....	42
	Gráfico 14 – Porcentagem de Migrantes por UF (PNAD 2001).....	43
	Gráfico 15 – Porcentagem de Migrantes após 1995 por UF: Viés de migração (PNAD 2001).....	44
	Gráfico 16 – Porcentagem de Migrantes por UF (Censo 2000).....	45
	Gráfico 17 – Porcentagem de Migrantes após 1995 por UF: viés de migração (Censo 2000).....	45
	Tabela 2 – Destino dos Jovens Migrantes após 1995 (PNAD 2001).....	46
	Tabela 3 – Destino dos Jovens Migrantes após 1995 (Censo 2000).....	46
4.2	Análise da Educação dos jovens no Brasil em 1995.....	46
	Gráfico 18 – Nível educacional dos jovens brasileiros em 1995.....	47
	Gráfico 19 – Porcentagem de Jovens por UF com 10 anos de estudo completos.....	47
	Gráfico 20 – Frequência à escola dos jovens em 1995.....	48
	Gráfico 21 – Viés Educacional.....	49
4.3	Análise da Educação dos pais da geração de 1977 e 1978 em 1982.....	50
	Gráfico 22 – Educação do Chefe do Domicílio em 1982 por UF.....	50
4.4	Sumário Estatístico.....	51
	Tabela 4 – Descrição das Variáveis (SAEB).....	52
	Tabela 5 – Descrição das Variáveis (PNAD, Censo e SAEB).....	52
	Figura 3 – Variáveis.....	53
5	RESULTADOS ECONOMÉTRICOS.....	53
	Tabela 6 – Proficiência em Matemática (1995) e Salários (2001_PNAD).....	54
	Tabela 7 – Proficiência em Português (1995) e Salários (2001_PNAD).....	56
	Tabela 8 – Proficiência em Matemática (1995) e Salários (2000_Censo).....	58
	Tabela 9 – Proficiência em Português (1995) e Salários (2000_Censo).....	59
5.1	Primeiro Estágio.....	61
	Tabela 10 – Primeiro Estágio: Função de Produção Educacional – Matemática.....	62
	Tabela 11 – Primeiro Estágio: Função de Produção Educacional – Português.....	63
5.2	Segundo Estágio.....	64
	Tabela 12 – Segundo Estágio: Proficiência em Matemática (1995) e Salários (2001_PNAD).....	65

Tabela 13 – Segundo Estágio: Proficiência em Português (1995) e Salários (2001_PNAD).....	67
Tabela 14 – Segundo Estágio: Proficiência em Matemática (1995) e Salários (2000_Censo).....	68
Tabela 15 – Segundo Estágio: Proficiência em Português (1995) e Salários (2000_Censo).....	70
5.3 Interações.....	72
Tabela 16 – Coeficiente de Determinação e Teste F de significância conjunta: Logaritmo da Proficiência em Matemática (1995)	73
Tabela 17 – Coeficiente de Determinação e Teste F de significância conjunta: Logaritmo da Proficiência em Português (1995).....	74
Tabela 18 – Coeficiente de Determinação e Teste F de significância conjunta: Logaritmo do Salário Real Horário (PNAD 2001)	75
Tabela 19 – Coeficiente de Determinação e Teste F de significância conjunta: Logaritmo do Salário Real Horário (Censo 2000).....	75
6 CONCLUSÃO.....	76
7 BIBLIOGRAFIA	78

1 INTRODUÇÃO

O processo de educação brasileiro pode ser caracterizado como lento (mesmo quando comparado a países com nível de desenvolvimento inferior ao nosso), bastante desigual e concentrado em uma parcela privilegiada da população. O debate em torno do peso da educação como fator explicativo da desigualdade de renda existente no Brasil é intenso, destacando-se a corrente que defende que a escolaridade é o principal fator explicativo causal da desigualdade, por gerar diferenças de produtividade entre os indivíduos que vão perdurar por todo seu ciclo de vida [Menezes-Filho (2001)]. Além disso, o capital humano é um dos principais determinantes da taxa de crescimento e do nível de bem-estar de um país.

Muitos estudos internacionais mostram que a qualidade da educação influencia positivamente os salários futuros dos indivíduos [Murnane et. al. (1995), Murphy e Peltzman (2004)], a probabilidade de continuação dos estudos [Rivkin (1995)] e o crescimento econômico do país [Bishop (1989), Hanushek e Kimko (2000)].

Desde a década de 1980, muitas mudanças institucionais foram implementadas no Brasil, incluindo diversas políticas educacionais, visando reduzir a desigualdade, ampliar o acesso às escolas e melhorar os índices de alfabetização. Resultados quantitativos de tais políticas vêm sendo observados, como uma aceleração recente do ritmo da evolução educacional no país. Embora tenham sido realizadas no Brasil muitas pesquisas sobre a qualidade do ensino, principalmente usando testes de proficiência, seus impactos sócio-econômicos ainda são desconhecidos.

Nesse trabalho analisamos a relação entre o desempenho escolar e os salários dos jovens brasileiros. Examinamos se a qualidade do ensino, medida pelas notas obtidas por uma geração nos exames de proficiência afeta os salários a serem recebidos por esta geração quando ela estiver no mercado de trabalho, cinco e seis anos depois. A partir de um modelo de pseudo-painel, corrigimos o problema de viés de seleção gerado pelas migrações e pelo alto nível educacional da amostra selecionada, através do modelo de Dahl (2001). Os determinantes da proficiência escolar, tais como “*background*” familiar, a infra-estrutura escolar e o perfil de docentes e diretores também são analisados, como possíveis determinantes da qualidade da educação.

Na segunda seção deste trabalho fazemos uma revisão aprofundada da bibliografia internacional sobre os impactos sócio-econômicos da qualidade da educação, medida pelo desempenho dos alunos em testes de proficiência, a fim de permitir a consolidação de uma base de conhecimento ampla e diversificada sobre o tema, abordando as diversas perspectivas de análise existentes. São recuperados os estudos semelhantes realizados nas últimas décadas em vários países. Na seção seguinte apresentamos detalhadamente o modelo de Roy (1951) de mobilidade e salários em que se discutem os efeitos da auto-seleção entre as diferentes ocupações. É um modelo de seleção baseado em vantagens comparativas que investiga os efeitos resultantes da distribuição de salários em diferentes ocupações. Este modelo é desenvolvido em Dahl (2001) que introduz algumas extensões. Ainda na terceira seção mostramos a metodologia econométrica desenvolvida nesse trabalho, destacando as características do modelo e as equações que estimamos. Na seção quatro destacamos as características e origem dos dados que utilizamos, e descrevemos como construímos as variáveis, apresentando gráficos estatísticos de algumas delas. Ainda nessa seção fazemos uma análise da migração e do nível educacional no Brasil para a amostra selecionada nesse trabalho e da educação dos pais da geração em análise. Na seção cinco apresentamos os resultados das estimações. Por fim, na seção seis destacamos as conclusões que obtivemos com o trabalho.

2 REVISÃO BIBLIOGRÁFICA

Em um esforço para melhorar a qualidade da escola, governos ao redor do mundo têm aumentado em grandes proporções os recursos destinados à educação. O interesse de estudiosos e políticos em melhorar a escola tem sido reforçado diretamente pelo reconhecimento da importância da formação do capital humano para indivíduos e sociedade. Muitas das motivações vêm de análises teóricas e empíricas sobre a relação entre a renda, a produtividade e o crescimento econômico e a quantidade de escolaridade do indivíduo, a *proxy* mais comum para nível de capital humano. Para muitos, entretanto, as iniciativas de políticas não deveriam focar em quantidade de escolaridade, mas ao invés disso, na sua qualidade [Hanushek (2003)].

Os argumentos centrais a respeito dos efeitos de escolaridade têm uma estrutura simples. Primeiro, existe uma relação direta entre escolaridade e salários. Dados estes retornos à educação, a discussão de política sugere a necessidade de se investir mais em capital humano, o que é traduzido diretamente em fornecer mais fundos para escola pública. O argumento para aumentar fundos geralmente não trata somente de aumentar os anos de estudo dos indivíduos, mas, além disso, em melhorar a qualidade dos anos de escolaridade existentes. O pressuposto é que maiores recursos vão melhorar a qualidade da escola, isto é, que há uma estreita relação entre os recursos e a qualidade. Essa relação, porém, é questionada em Hanushek (2003) como veremos adiante.

Há muitas evidências internacionais de que a qualidade da educação, medida por testes de proficiência, está relacionada com os salários individuais, a produtividade e o crescimento econômico, mesmo após o controle pela quantidade de escolaridade [Bishop (1989), O'Neill (1990), Murnane *et al.* (1995), Boissiere *et al.* (1985)]. Segundo Murnane *et al.* (1995), existe uma relação entre as notas dos testes e os salários, que vem aumentando com o passar do tempo. Os autores examinam se o conhecimento cognitivo básico está se tornando mais importante na determinação dos salários na economia. A questão central focada é como o conhecimento em matemática dos estudantes afeta seus salários aos 24 anos quando estão trabalhando, seis anos após o término de seus cursos. A análise é feita separadamente para os homens e para as mulheres, avaliando tal questão para os estudantes que terminaram o curso em 1972, e novamente para aqueles que completaram o curso em 1980. O artigo mostra que a habilidade cognitiva básica teve um maior impacto nos salários dos indivíduos em 1986 do que em 1978. A introdução do resultado da proficiência no modelo básico resulta em um declínio no coeficiente de anos de escolaridade de 41% em 1972 e 52% em 1980 para os homens. Para as mulheres as reduções foram de 31% e 43%, respectivamente. O aspecto impressionante do achado desses autores é que adicionando essa nota ao modelo elimina completamente o aumento observado do prêmio de salário das mulheres por freqüentar a universidade entre 1978 e 1986, que passa a ser explicado pelo aumento no retorno ao conhecimento cognitivo [Murnane *et al.* (1995)].

Existem ainda evidências de que os estudantes que são mais bem avaliados nos resultados de testes de proficiência tendem a ir mais longe nos estudos. Por isto, Rivkin (1995), Hanushek (1996) e Bishop (1991) concluem que as notas conseguidas pelos indivíduos são altamente correlacionadas com a freqüência escolar futura e que a conclusão da escola é significativamente relacionada à melhores resultados em proficiência.

Corroborando os estudos anteriores, Boissiere *et al.* (1985) procura distinguir a influência nos salários das habilidades cognitivas, da habilidade nativa, e dos anos de escolaridade. Utilizando dois conjuntos comparáveis de micro dados, do Quênia e da Tanzânia, os autores estimam o efeito do conhecimento cognitivo (representado pelo desempenho em exames de proficiência), da habilidade e dos anos de escolaridade nos salários. Os autores concluem que com a introdução da variável de conhecimento cognitivo, o prêmio por anos de estudo declina cerca sessenta por cento, e na Tanzânia não é mais significativamente diferente de zero. Em nenhum dos dois países a influência da habilidade nos salários é grande ou significativa. Em contraste, em ambos o coeficiente do teste de proficiência é positivo, significativo, e grande relativamente ao de habilidade.

O trabalho de Bishop (1989) destaca o grande declínio da nota média dos testes de proficiência ocorrido entre 1967 e 1980 nos Estados Unidos, que gerou uma redução da produtividade, que, por sua vez, gerou custos sociais. Existem especulações de que o declínio da nota do SAT (Scholastic Aptitude Test) pode sinalizar uma grande queda na qualidade dos jovens ingressantes da força de trabalho e que isso seja parcialmente responsável pela redução do crescimento da produtividade. Ao analisar a relação entre a nota do teste e o crescimento da produtividade, Bishop (1989) conclui que o momento de declínio do resultado do teste (começando em 1967 e terminando em 1980) é extraordinariamente coincidente com o declínio do crescimento da produtividade. A redução na nota da proficiência começou a ter efeito no crescimento da produtividade em meados da década de 70. O autor estima que se essa nota tivesse continuado a crescer depois de 1967 (a taxa que prevaleceu no quarto de século anterior), a qualidade da força de trabalho seria 2,9 por cento maior em 1987 e o PIB desse ano de US\$86 bilhões a mais. O cálculo do custo social a ser acumulado em 2010 devido a isso seria, a valor presente de 1987, de US\$3,2 trilhões.

Assim, além do indivíduo, a sociedade também é influenciada pela qualidade educacional, em termos de ganhos de produtividade e de crescimento econômico. Hanushek e Kimko (2000), utilizando informações de exames internacionais de matemática e ciências, demonstram que as diferenças na qualidade das escolas têm grandes impactos nas taxas de produtividade e de crescimento nacional. Os autores mostram que a qualidade do ensino tende a ser mais importante para o crescimento e o bem-estar de um país do que a média de anos de estudo da sua população.

Os resultados de uma das regressões para os países, que descreve o crescimento médio real per capita (GPD) entre 1960 e 1990, mostram que a quantidade de escolaridade tem um forte e positivo impacto no crescimento quando esse é mensurado na ausência das medidas de

qualidade. Entretanto, ao adicionar as medidas de qualidade da força de trabalho os autores obtêm uma relação muito forte entre a qualidade e a taxa de crescimento per capita, dobrando o valor do coeficiente de determinação do modelo (R^2). De acordo com os resultados obtidos, o aumento de um desvio padrão da qualidade da força de trabalho, aumenta a taxa de crescimento per capita em cerca de 1,4 ponto percentual por ano. Em contraste, um desvio padrão a mais na quantidade de escolaridade está associado a um aumento de apenas um quarto de ponto percentual no crescimento. Além disso, o efeito da quantidade de escolaridade cai muito com a inclusão das medidas de desempenho. Esses resultados fornecem forte sustentação para a importância da diferença na qualidade da força de trabalho medida pelo conhecimento em matemática e em ciências [Hanushek e Kimko (2000)].

Murphy e Peltzman (2004) fazem uma análise para os Estados Unidos semelhante à realizada neste trabalho para o Brasil. Os autores examinam como a qualidade da educação recebida pelas crianças afeta seu desempenho quando elas ingressam no mercado de trabalho. Eles observam o mercado de trabalho dos Estados Unidos desde 1970 até meados da década de 1990, período no qual a demanda e os retornos à habilidade aumentaram. Os autores analisam como a qualidade da escola, medida por testes de proficiência, relaciona-se com os salários dos jovens que entram no mercado de trabalho, com a qualidade do trabalho e com a continuidade escolar. Segundo seus resultados, um desvio padrão de deterioração da nota relativa de um estado corresponde a uma redução de 3% no salário médio dos jovens ingressantes na força de trabalho. Além disso, a qualidade da educação foi ao menos tão importante nas variações salariais ocorridas quanto as mudanças no mercado de trabalho, e também afeta a qualidade do trabalho obtido por estes jovens. Os autores concluem que o efeito do desempenho escolar no salário é resultado tanto da maior qualidade do trabalho como dos maiores salários pagos no mesmo emprego, e mostraram que melhores resultados escolares aumentam as matrículas nas Universidades [Murphy e Peltzman (2004)].

Os estudos recentes sobre os retornos a educação no Brasil enfatizam o impacto da escolaridade, medida em anos de estudo, sobre os salários. Porém, muitas pesquisas internacionais têm reconhecido a superioridade da qualidade desses anos de estudo, mensurada por testes de proficiência, sobre a quantidade para explicar os salários dos estudantes quando ingressam no mercado de trabalho, assim como a produtividade dos países e das regiões e suas diferenças. O que se tem observado no país é que a educação avança, mas não em termos qualitativos. A divulgação dos resultados do Sistema Nacional de Avaliação do Ensino Básico (SAEB) documenta uma queda no desempenho escolar entre 1995 e 1999

(Ministério da Educação, 2001). Os resultados divulgados das provas do SAEB realizadas em 2001 mostram que a tendência de queda continua [Menezes-Filho (2004)].

O próprio MEC reconhece que há problemas de qualidade na educação básica brasileira. Segundo o levantamento feito pelo órgão em 2003, 55% dos alunos que concluem a quarta série do ensino fundamental tem desempenho em leitura considerado crítico ou muito crítico, sendo que no Nordeste e Norte esse valor chega a 70% e 66%, respectivamente, enquanto para o Sudeste e o Sul as proporções são de 44% e 47%. Para os alunos que concluem a terceira série do ensino médio esse fraco desempenho é observado em quase 40% deles. Novamente destacam-se as disparidades regionais, no Sul apenas 29% tem nível de leitura crítico ou muito crítico, enquanto no Norte e no Nordeste essa porcentagem é bem maior, 51% e 48%, respectivamente [SAEB (2004)].

Nos últimos anos foram implementadas no Brasil políticas visando aumentar o acesso de crianças às escolas, reduzir as reprovações e a evasão escolar. Entretanto, os resultados positivos de tais medidas quantitativas são questionados por pesquisas que avaliam a qualidade do ensino. Este trabalho analisa o impacto da qualidade da educação sobre os salários dos jovens trabalhadores. O objetivo é avaliar se políticas que visam a melhoria da qualidade dos anos de estudos recebidos pelos jovens na escola têm impactos sociais e econômicos positivos para o Brasil.

3 METODOLOGIA ECONOMÉTRICA

A análise econométrica que realizamos nesse trabalho apresenta duas peculiaridades: pseudo-painel e viés de seleção. A literatura sobre pseudo-painel foi introduzida por Browning *et al.* (1985) e é utilizada por pesquisadores que não dispõem de dados em painel, mas de vários cortes transversais, independentes entre si, nos quais indivíduos diferentes são entrevistados a cada período. O objetivo dessa técnica é superar as limitações dos cortes transversais, aproveitando as vantagens desses cortes repetidos com relação aos dados de painel. As limitações decorrem do fato do pesquisador não dispor de valores defasados das variáveis, o que tornaria impossível, em princípio, o controle dos efeitos específicos e a identificação de modelos dinâmicos. As vantagens dizem respeito a maior cobertura temporal desse tipo de dado, ao fato de não ser suscetível a problemas de atrito e diminuir os erros de medida, por trabalhar com dados mais agregados [Menezes-Filho (2001b)].

Neste trabalho, o pseudo-painel consiste em um agrupamento dos indivíduos por células de acordo com a coorte de nascimento. Analisamos as coortes nascidas em 1977 e 1978 em três momentos da vida: em 1982 (quando essa geração tinha 4 e 5 anos de idade), em 1995 (ao final do ensino médio, aos 17 e 18 anos de idade), e em 2000 ou 2001 (aos 22 e 23, ou 23 e 24 anos de idade, quando esta geração está no mercado de trabalho). Esse agrupamento se dá através da média da variável dependente (logaritmo dos salários) e dos regressores para cada momento do tempo. Assim, ao invés de seguir os mesmos indivíduos, são acompanhadas as gerações ao longo do tempo através das coortes.

O problema de viés de seleção aparece duas vezes. A primeira delas porque todas as pessoas que compõe a amostra do SAEB, isto é, que fizeram as provas de proficiência ao final do ensino médio, têm 10 anos de estudo e freqüentam a escola. São alunos da 3^a. série do ensino médio que fizeram a prova de proficiência quando cursavam o último ano do ensino médio. Isto significa que a amostra é composta por pessoas que possuem um nível educacional elevado. O segundo viés é o de migração. São excluídos das amostras de 2000 e 2001 os jovens que migraram após os 17 e 18 anos de idade, a fim de garantir que esses fizeram o exame de proficiência aos 17 e 18 anos de idade no mesmo estado em que moram e trabalham cinco e seis anos depois. Assim, a amostra utilizada nesse trabalho não é obtida aleatoriamente, mas é uma amostra selecionada. Essas auto-seleções da amostra geram dois viés de seleção os quais corrigimos aplicando o modelo clássico proposto por Roy (1951) e testado em Dahl (2002).

3.1 MODELO DE ROY POR DAHL

Para captar o impacto da auto-seleção nos retornos estimados no mercado de trabalho dos Estados Unidos, o trabalho de Dahl (2002) desenvolve o modelo de Roy (1951) de mobilidade e salários em que os trabalhadores escolhem em qual dos 50 estados americanos mais o Distrito de Columbia morar e trabalhar. O mercado de trabalho americano caracteriza-se pela ampla variação dos retornos estimados para um nível educacional entre os estados, e pela alta taxa de mobilidade dos trabalhadores, isto é, as pessoas não trabalham no mesmo estado em que nasceram. A migração auto-selecionada apresenta uma explicação de porque os retornos à educação não são iguais entre os estados: se os trabalhadores escolhem onde morar e trabalhar baseados em vantagens comparativas, então os retornos estimados à educação em qualquer estado devem ser viesados para cima ou para baixo [Dahl, (2002)].

Dahl aplica o modelo de Roy de mercado múltiplo de mobilidade e salários para entender os efeitos da migração auto-selecionada e obter estimativas não viesadas do retorno à escolaridade. Ao invés dos trabalhadores escolherem entre as ocupações como em Roy, o artigo de Dahl formula um modelo em que indivíduos escolhem em quais dos 51 estados americanos vão morar e trabalhar. Como no modelo de Roy, as diferentes áreas geográficas são modeladas tendo salários e benefícios distintos que atraem os trabalhadores com diferentes níveis de escolaridade.

Assim, a migração auto-selecionada gera um retorno observado à educação para residentes atuais de um estado diferente do retorno que um indivíduo escolhido aleatoriamente esperaria ganhar. Para corrigir esse viés de seleção amostral em um modelo de múltiplas escolhas, Dahl utiliza uma nova metodologia semi-paramétrica.

3.1.1 Uma extensão do Modelo de Roy: Um Modelo de Mobilidade e Salários

O artigo de Roy de 1951 discute os efeitos da auto-seleção entre diferentes ocupações. É um modelo de seleção baseado em vantagens comparativas que investiga os efeitos resultantes da distribuição de salários em diferentes ocupações. Dahl desenvolve o modelo de Roy para a escolha individual de onde morar e trabalhar. Os salários em diferentes regiões variam pelo nível de escolaridade, e cada indivíduo segue o caminho da transição que maximiza sua utilidade. A busca de vantagens comparativas gera retornos observados à educação em uma área diferente da verdadeira média populacional.

O modelo de Roy considera apenas duas ocupações ou setores, é baseado na maximização da renda e assume uma distribuição log-normal da habilidade latente. Dahl introduz três extensões para esse modelo: considera múltiplos mercados ou setores, que as escolhas são baseadas na maximização da utilidade, e que a distribuição da habilidade latente não é especificada.

O modelo de Dahl considera um país com N áreas geográficas distintas e que os indivíduos vivem por dois períodos. No primeiro período os indivíduos nascem e não trabalham, enquanto no segundo período os indivíduos trabalham. Os indivíduos são aleatoriamente fixados na área geográfica em que nasceram no primeiro período, e escolhem onde vão morar e trabalhar no segundo período de suas vidas. O estado de nascimento caracteriza-se pelo local onde os indivíduos nasceram, e o estado de residência é onde os indivíduos escolhem trabalhar. Na ausência de migrações, cada área tem a mesma distribuição

de habilidades individuais, a migração auto-selecionada altera potencialmente a distribuição de habilidades entre os estados.

O modelo considera que os indivíduos fizeram suas decisões de migração e começaram a trabalhar. A função de salário populacional para o indivíduo i que trabalha no estado k é dada por:

$$(1) \quad y_{ik} = a_k + x_i' d_k + s_i b_k + u_{ik}, \quad k = 1, \dots, N$$

y_{ik} : logaritmo do salário

a_k : constante específica do estado

x_i : vetor de características individuais

s_i : medida do nível de escolaridade

u_{ik} : termo de erro

O problema de estimação é que o logaritmo do salário é observado apenas na região onde a pessoa se encontra. Seu salário nas outras regiões é desconhecido, o que é um problema, pois a decisão de migrar de uma região para outra, que faz com que o salário seja observado, depende dos salários nas duas regiões. Em uma amostra auto-selecionada do estado k , o termo de erro u_{ik} não tem, necessariamente, média zero condicional em x_i e s_i , e a regressão por MQO gera uma estimação viesada de d_k e b_k .

O artigo de Dahl mensura o retorno à educação nos 51 estados corrigidos pelo viés de seleção da migração induzida. A migração entre os estados é baseada na maximização da utilidade em relação à decisão de migrar da região j para a k . O autor assume que a utilidade do indivíduo i , nascido no estado j , considerando a migração para o estado k , consiste de uma função aditiva de dois fatores: os salários e as preferências pessoais específicas.

$$(2) \quad V_{ijk} = y_{ik} + t_{ijk}, \quad k = 1, \dots, N$$

V_{ijk} : índices de utilidade

y_{ik} : logaritmo dos salários

t_{ijk} : vetor de preferências de mudar do estado j para o estado k

O vetor t_{ijk} representa os fatores que não determinam os salários e que entram na função de utilidade, tais como qualquer custo de deslocamento, as diferenças de comodidades entre os estados, qualquer outro fator que não explica os salários, e os benefícios associados ao deslocamento de um estado para outro.

O desvio dos ganhos individuais de trabalhar no estado k da média da população inteira (incluindo indivíduos que atualmente não trabalham no estado k) é dado por:

$$(3) \quad y_{ik} - E[y_{ik} | x_i, s_i] = u_{ik}, \quad k = 1, \dots, N$$

De forma similar, os desvios das preferências individuais de mudar de estado j para o estado k da média populacional são dados por:

$$(4) \quad t_{ijk} - E[t_{ijk} | z_i] = w_{ijk}, \quad k = 1, \dots, N$$

z_i : vetor de características individuais

w_{ijk} : termo de erro do desvio individual da preferência média

A função V_{ijk} é chamada de função de sub-utilidade e é escrita em termos da média populacional e de um componente de erro específico para o indivíduo.

$$(5) \quad V_{ijk} = V_{jk} + e_{ijk}, \quad k = 1, \dots, N$$

em que $V_{jk} = E[y_{ik} | x_i, s_i] + E[t_{ijk} | z_i]$ e $e_{ijk} = u_{ik} + w_{ijk}$.

Como os indivíduos seguem a trajetória de migração que maximiza sua utilidade, o indivíduo i escolhe migrar do estado j para o estado k de acordo com a seguinte regra:

$$M_{ijk} = 1 \text{ se e, somente se, } V_{ijk} = \max(V_{ij1}, \dots, V_{ijN}),$$

$M_{ijk} = 0$, caso contrário, em que M_{ijk} é o indicador de se o indivíduo i migra do estado j para o estado k .

A equação de seleção pode ser escrita como:

$$(6) \quad M_{ijk} = 1 \text{ se, e somente se, } V_{jk} + e_{ijk} \geq V_{jm} + e_{ijm}, \forall m,$$

$M_{ijk} = 0$, caso contrário.

Isto quer dizer que se o indivíduo é observado na região k isto reflete o fato de que a utilidade obtida de migrar de j para k é estritamente maior que a utilidade de migrar para outras regiões. A utilidade depende da trajetória específica da migração do estado j para o k , isto é, a utilidade de um indivíduo depende não apenas do estado de residência, mas também do estado de nascimento.

A regra de seleção é dada por:

$$(7) \quad y_{ik} \text{ observado se, e somente se, } M_{ijk} = 1$$

Assim, os salários são observados em uma escolha de maximização da utilidade individual. Os salários são observados apenas se as N equações de seleção em (6) são satisfeitas simultaneamente. As equações (1) a (7) descrevem um modelo de Roy estendido de salários e mobilidade.

Os indivíduos que vivem no estado k não são uma amostra aleatória da população, e em geral,

$$(8) \quad E[u_{ik} | y_{ik} \text{ observado}] = E[u_{ik} | M_{ijk} = 1]$$

$$E[u_{ik} | y_{ik} \text{ observado}] = E[u_{ik} | e_{ijm} - e_{ijk} \leq V_{jk} - V_{jm}, \forall m]$$

$$E[u_{ik} | y_{ik} \text{ observado}] \neq 0, \text{ em que } E[u_{ik} | M_{ijk} = 1] \text{ é o viés de seleção da observação } i.$$

Se essa expectativa condicional é correlacionada com x_i ou s_i , a regressão por MQO de y_{ik} em x_i e s_i vai resultar numa estimação viesada. A direção e o tamanho do viés para um indivíduo dependem da distribuição conjunta de u_{ik} e o termo de erro das N equações de migração, $e_{ij1} - e_{ijk}, \dots, e_{ijN} - e_{ijk}$. Como $e_{ijk} = u_{ik} + w_{ijk}$, o viés depende da correlação dos salários residuais e das preferências residuais.

3.1.2 Modelo de Viés de Seleção com Múltiplas Escolhas

Dahl inicia modelando a distribuição conjunta dos termos de erro na equação de salários e na equação de seleção. A dificuldade em explicar a correlação dos termos de erro das N equações de seleção com o termo de erro da equação de salário faz com que Dahl utilize a sugestão de Lee (1983)¹ de reduzir a dimensão do problema pela reformulação das N equações de seleção em (6) em termos das estatísticas de ordem.

Combinando as equações (6) e (7), a regra de seleção do estado k torna-se:

$$y_{ik} \text{ observado se, e somente se, } (V_{j1} - V_{jk} + e_{ij1} - e_{ijk}, \dots, V_{jN} - V_{jk} + e_{ijN} - e_{ijk})' \leq 0$$

A expressão equivalente é:

$$(9) \quad y_{ik} \text{ observado se, e somente se, } \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) \leq 0, \text{ em que } \max_m (\bullet)$$

indica o máximo para m .

Assim, qualquer viés de seleção em y_{ik} é direcionado pelo evento em que o máximo da coleção de variáveis aleatórias, $V_{j1} - V_{jk} + e_{ij1} - e_{ijk}, \dots, V_{jN} - V_{jk} + e_{ijN} - e_{ijk}$, é menor ou igual a zero.

A hipótese implícita na transformação de Lee é que a distribuição conjunta do termo de erro na equação de resultado e a estatística de ordem máxima não dependem das diferenças entre as funções de sub-utilidade. A transformação de Lee do máximo é central porque a mesma transformação é aplicada apesar dos valores específicos de $V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}$. A hipótese implícita na aproximação de Lee para um modelo de escolha multivariada é:

¹ Lee, L. (1993). 'Generalized Econometric Model with Selectivity' *Econometrica*, 51, 507-512.

$$(A-1) \quad g_k \left(u_{ik}, \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) \right) \text{ não depende de } V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}.$$

Dahl propôs uma aproximação alternativa motivada pela observação de que no modelo de seleção de índice único o viés de seleção pode ser escrito como uma função de probabilidade de seleção, dadas as covariáveis.

A formulação de mobilidade e salários nas equações (1) e (6) implica que as equações de salários podem ser reescritas como um modelo parcialmente linear de índice múltiplo.

$$(10) \quad y_{ik} = a_k + x_i' d_k + s_i b_k + \sum_{j=1}^N [M_{ijk} \times y_{jk} (V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk})] + u_{ik}, \quad (k=1, \dots, N)$$

$$\text{em que } y_{jk}(\bullet) = E[u_{ik} \mid V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}]$$

u_{ik} : termo de erro com média zero na amostra condicional para o estado k .

A equação (10) é chamada de modelo de índice múltiplo, porque as funções y_{jk} para cada estado de nascimento j são funções desconhecidas dos índices múltiplos $V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}$. A dimensão das funções controle pode ser reduzida usando a modificação proposta pela aproximação de Lee.

Para obter vantagem da idéia de Lee numa estrutura semi-paramétrica, Dahl faz a seguinte hipótese de suficiência de índice:

$$(A-2) \quad \begin{aligned} & g_k \left(u_{ik}, \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk} \mid V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}) \right) \\ & = g_k \left(u_{ik}, \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk} \mid p_{ijk}) \right) \end{aligned}$$

p_{ijk} : probabilidade do indivíduo i mover do estado j para o estado k , dado o vetor $V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}$. A equivalência em (A-2) assume que $p_{ijk} = p_{ijk}(V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk})$ que esgota toda a informação sobre como $V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}$ influencia a distribuição conjunta de u_{jk} e $\max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ contida na amostra. Isto é, a distribuição condicional de u_{jk} e $\max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ pode depender das variáveis condicionais apenas através do índice único p_{ijk} , que é a probabilidade da primeira melhor escolha individual de migração, uma escolha que é observada desde que o pesquisador saiba onde um indivíduo escolhe morar e trabalhar. Essa probabilidade de migração associada com a estatística de ordem máxima é escrita como:

$$(11) \quad \begin{aligned} p_{ijk} & = \Pr(M_{ijk} = 1 \mid V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}) \\ p_{ijk} & = \Pr(V_{jk} + e_{ijk} \geq V_{jm} + e_{ijm}, \forall m) \end{aligned}$$

A função sub-utilidade tem que ser contabilizada para obter uma estimativa de p_{ijk} , em que o vetor $V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}$ determina a escolha de migração individual.

Usando a hipótese (A-2), a equação de salários é escrita como índice único no modelo parcialmente linear:

$$(12) \quad y_{ik} = a_k + x_i' d_k + s_i b_k + \sum_{j=1}^N [M_{ijk} \times I_{jk}(p_{ijk})] + w_{ik}, \quad (k=1, \dots, N)$$

em que para cada estado j , $I_{jk}(\bullet)$ é uma função desconhecida do índice único p_{ijk} , e w_{ik} é o termo de erro. Dahl se refere as I_{jk} 's como as funções de correção de seleção para o estado k . Por construção, o termo de erro w_{ik} tem média zero dada a probabilidade de migração e o fato de que os salários são observados em um estado:

$$E[w_{ik} | x_i, s_i, p_{ijk}, M_{ijk} = 1] = 0, \quad (k=1, \dots, N)$$

$$\text{Assim, se a hipótese (A-2) é mantida: } y_{jk}(V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}) = I_{jk}(p_{ijk})$$

O uso de p_{ijk} na função de correção do modelo de múltipla escolha é interpretado pelo fato de que, sujeito a condição de invertibilidade, diz que as probabilidades de migração múltiplas, p_{ij1}, \dots, p_{ijN} , contém a mesma informação que as diferenças entre as funções de sub-utilidade, $V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}$. Isso implica que as equações de salários podem ser reescritas como um modelo parcialmente linear de índice múltiplo que depende de todas as N probabilidades de migração.

$$(13) \quad y_{ik} = a_k + x_i' d_k + s_i b_k + \sum_{j=1}^N [M_{ijk} \times m_{jk}(p_{ij1}, \dots, p_{ijN})] + u_{ik}, \quad (k=1, \dots, N)$$

$$\text{em que } m_{jk}(\bullet) = E[u_{ik} | p_{ij1}, \dots, p_{ijN}] = E[u_{ik} | V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk}].$$

A hipótese (A-2) simplifica essa equivalência assumindo que apenas a probabilidade de escolha de maximização da utilidade resolve a parametrização da distribuição conjunta g_{jk} . Portanto, (A-2) também pode ser pensada como uma restrição exclusiva que requer que a distribuição de u_{jk} e $\max_m (V_{jm} - V_{jk} + e_{jm} - e_{ijk})$ dada p_{ij1}, \dots, p_{ijN} seja a mesma que dada p_{ijk} .

O relaxamento de (A-2) permite outras probabilidades, que ao lado da probabilidade de primeira melhor escolha, também influencia a distribuição conjunta g_{jk} . Considerando \vec{q} o subconjunto escolhido de todo conjunto de probabilidades de migração $\{p_{ij1}, \dots, p_{ijN}\}$, uma hipótese menos restritiva é escrita como

$$(A-3) \quad \begin{aligned} & g_{jk} \left(u_{ik}, \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) \mid V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk} \right) \\ & = g_{jk} \left(u_{ik}, \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) \mid p_{ijk}, \vec{q} \right) \end{aligned}$$

Essa extensão permite às equações de salários serem escritas como modelos parcialmente lineares de índice múltiplo, em que a correção do viés é uma função não conhecida da primeira melhor probabilidade de migração mais umas poucas outras probabilidades escolhidas. A dificuldade é decidir que outras probabilidades são importantes na parametrização da distribuição conjunta g_{jk} , sendo que apenas um seletor número de probabilidades podem ser incluídas para que a dimensionalidade não deixe a estimação inviável. Na aplicação do artigo são sugeridas algumas probabilidades: a probabilidade de retenção, isto é, a probabilidade da pessoa que nasceu no estado j escolher permanecer no estado j , a alta probabilidade de previsão de um indivíduo, excluindo a probabilidade de retenção, $\max_m (p_{ijm}), m \neq j$, e as probabilidades de migração de estados geograficamente perto do estado de nascimento do indivíduo.

Dahl coloca a probabilidade de retenção como um outro termo na função de correção. Assim, a hipótese distribucional fica,

$$(A-4) \quad \begin{aligned} & g_{jk} \left(u_{ik}, \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) \mid V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk} \right) \\ & = g_{jk} \left(u_{ik}, \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) \mid p_{ijk}, p_{ijj} \right) \end{aligned}$$

que implica que as equações de salários podem ser escritas como

$$(14) \quad y_{ik} = a_k + x'_i d_k + s_i b_k + \sum_{j=1}^N \{ M_{ijk} \times I_{jk}^* (p_{ijk}, p_{ijj}) \} + w_{ik}^*, \quad (k=1, \dots, N)$$

Os termos de correção nas equações de salários para migrantes são funções desconhecidas de duas probabilidades p_{ijj} e p_{ijk} . Sendo que, para os não migrantes os termos de correção são uma função de uma única probabilidade, p_{ikk} , desde que $j=k$ para os indivíduos que não mudam de seus estados de nascimento.

3.1.3 Implementação das Escolhas

A hipótese de suficiência de índice reduz muito a dimensão das funções de correção de seleção.

3.1.3.1 Reduzindo o número de funções de correção

No modelo de Dahl, existem 51 funções de correção diferentes para cada estado de residência k , e uma função de correção diferente para cada possível estado de nascimento j . A hipótese (A-4) diminui a dimensão de cada uma dessas 51 funções de correção formando uma função de todas as 51 probabilidades para as 51 correções que são uma função da primeira melhor probabilidade de migração, p_{ijk} , e da probabilidade de retenção, p_{ijj} .

Uma dessas funções de correção é I_{kk}^* para não migrantes. Essa função corrige o viés de seleção dos indivíduos que nasceram no estado k e decidiram permanecer no estado k . As outras 50 funções, I_{jk}^* , ($j = 1, 2, \dots, 51, j \neq k$), corrigem na equação de salários o viés de seleção do imigrante do estado k para um dos outros 50 estados. Embora tenham 51 funções controle, o problema da dimensão é eliminado pela hipótese de suficiência de índice. Isso porque a taxa de convergência de uma função controle não paramétrica não é afetada pelo número de funções de correção incluídas na regressão.

Pelo modelo de Roy, para poucos setores, uma função de correção separada poderia ser incluída para cada setor de origem j . Entretanto, na aplicação de Dahl, estimar 51 diferentes funções para cada equação de regressão é incompatível. Em geral, a função de correção na equação (14) depende de uma distribuição conjunta diferente g_{jk} , para cada estado de origem j . Para reduzir o número de funções de correção que entram na equação de salários dos estados, Dahl assume,

$$(15) \quad g_{ik} = g_k, \forall j \neq k$$

A equação (15) restringe a distribuição para migrantes. Para um estado receptor k , a distribuição conjunta tem que ser a mesma para todos os possíveis estados de origem j (em que $j \neq k$). Essa equação implica que $I_{jk}^* = I_k^*$ para todos os estados de origem j diferentes de k , o que ajuda a identificar os coeficientes na equação de salários utilizando apenas duas funções de correção, uma para não migrantes e uma para migrantes, que entram na equação de salários dos estados apesar das N funções diferentes.

3.1.3.2 Usando Células de Fluxo Migratório para Possibilidades de Seleção.

Simplificando a estimação das probabilidades de migração, Dahl assume que os salários médios e as preferências médias são os mesmos para tipos de pessoa similares. Ele

supõe um vetor de variáveis contendo todos os atributos relevantes sobre um tipo de pessoa, tal que os indivíduos com os mesmos valores para essas variáveis de tipo de pessoa são identicamente afetados pelas diferenças estado por estado na função de sub-utilidade. As vantagens comparativas motivam essa aproximação, a previsão é que os indivíduos com diferentes habilidades e características vão seguir diferentes caminhos de migração na média.

Se o vetor descrevendo um tipo de indivíduo é composto apenas por variáveis discretas (tais como idade, escolaridade, raça, sexo, matrimonial, presença de crianças em casa), os indivíduos similares podem então ser agrupados por células. Com essa formulação, as funções de sub-utilidade dependem apenas da célula do indivíduo. A probabilidade de migração de um indivíduo pertencente à célula é:

$$(16) \quad p_{ijk} = \Pr(M_{ijk} = 1 | V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk})$$

$$p_{ijk} = \Pr(M_{ijk} = 1 | cell)$$

Uma expressão similar existe para a probabilidade de retenção, p_{ij} . Uma probabilidade de migração individual é simplesmente a fração de indivíduos na mesma célula que migraram de j para k . A vantagem de agrupar os indivíduos em células é que a forma da função de sub-utilidade latente não é especificada apropriadamente se indivíduos definidos similares têm as mesmas preferências e os mesmos salários em média. Adicionalmente, devido as células, a estimação da probabilidade de migração não requer hipóteses sobre os termos de erro, $e_{ij1} - e_{ijk}, \dots, e_{ijN} - e_{ijk}$.

Nesse trabalho as auto-seleções dos indivíduos ocorrem, pois as pessoas se auto-selecionam para estudar mais ou menos, assim como para migrar ou não migrar entre as Unidades da Federação brasileiras. Essas auto-seleções geram mudanças na distribuição de rendimentos. O fato de os migrantes mudarem de região por decisão própria, comparando os rendimentos potenciais, afeta bastante as estimativas de retorno à educação nas diferentes regiões. A auto-seleção de ser um migrante gera um viés de seleção nas estimativas de retorno à educação obtidas com uma amostra aleatória de pessoas que habitam uma região [Bastos e Ribeiro, (2004)]. O mesmo vale para o viés gerado pelos diferenciais educacionais entre os jovens, dados os elevados retornos à educação medida por anos de estudo como ressaltado pelas literaturas nacional e internacional, o chamado “efeito diploma”.

O método aplicado por Dahl (2002) utiliza para corrigir o viés de seleção gerado pela migração entre os estados americanos uma variável que representa a porcentagem de migrantes em cada estado. O autor introduz essa variável ao modelo na forma linear, quadrática e ao cubo. Para corrigirmos o viés de seleção desse trabalho são usadas duas

variáveis: uma que indica a porcentagem de jovens das coortes de 1977 e 1978 em 2000/2001 que não migraram de 1995 em diante, e uma outra variável que indica a porcentagem de pessoas dessas coortes com exatamente 10 anos de estudo que frequentavam a escola em 1995, ano em que fizeram a prova de proficiência do SAEB, pois os indivíduos que fazem tal prova estão terminando a 3^a. série do ensino médio. Isto é, essa variável indica a porcentagem de pessoas em condições de fazer o exame de proficiência.

3.2 O MODELO ESTIMADO

Estimamos o modelo de pseudo-painel através do método de Mínimos Quadrados, tendo como variável dependente o logaritmo dos salários e como variáveis explicativas o logaritmo da proficiência, as variáveis de correção do viés de seleção e as variáveis relacionadas às diversas características dos indivíduos².

$$LWRH_{c,t} = \alpha + \beta_1 Profic_{c,95} + \beta_2 ViésMigração_{c,t} + \beta_3 ViésEducacional_{c,95} + \beta_4 \chi_c + u_c \quad (1)$$

Em que c representa as células ($c = 1, \dots, 108$), χ_c são as variáveis relacionadas às diversas características dos tipos de indivíduos e t os anos de 2000 e 2001.

Primeiramente vamos analisar o efeito global dos testes de proficiência sobre os salários futuros, como apresentado na equação (1). Em seguida, procuraremos analisar se o impacto da qualidade da educação sobre o salário é oriundo do componente observável da qualidade (medida pelos insumos educacionais) ou pelo seu componente não-observável (habilidade do aluno, insumos não mensurados, etc).

Para alcançar este objetivo realizamos a estimação em dois estágios. No primeiro estágio são investigados os determinantes do desempenho escolar. No segundo estágio, são incluídos tanto o desempenho escolar bruto como os resíduos da regressão do primeiro estágio. Desta forma os resíduos vão capturar a parcela do desempenho escolar não explicada no modelo enquanto que a nota do SAEB vai capturar o valor previsto pela regressão.

Assim, no primeiro estágio fazemos as estimações por Mínimos Quadrados Ordinários das notas obtidas pelos alunos das coortes considerados nos exames de proficiência de 1995 em função das características dos mesmos, dos seus professores e diretores, e das escolas.

² O programa Stata foi usado para rodar todas as regressões descritas neste artigo.

Primeiro estágio:

$$Profic_{i,d} = \gamma + \theta_1 A_{i,t} + \theta_2 P_{i,t} + \theta_3 D_{i,t} + \theta_4 E_{i,t} + v_{i,t} \quad \Rightarrow \quad RESID_{i,d} \quad (2)$$

Em que i representa os indivíduos, d indica as disciplinas (matemática e português), $RESID$ são os resíduos da regressão de proficiência em função das características dos alunos, A , professores, P , diretores, D , e escolas E .

No estágio seguinte, adicionamos à equação (1) o resíduo obtido no primeiro estágio, agregados pelos vetores características, e estimamos por Mínimos Quadrados um modelo de pseudo-painel em que regredimos o logaritmo do salário em função do logaritmo da proficiência, das variáveis de correção do viés de seleção, das variáveis relacionadas às diversas características dos indivíduos e do resíduo da estimação do primeiro estágio.

Segundo estágio:

$$LWRH_{c,t} = \alpha + \beta_1 Profic_{c,95} + \beta_2 ViésMigração_{c,t} + \beta_3 ViésEducac_{c,95} + \beta_4 \chi_c + \beta_5 Resid_{c,95} + u_c \quad (3)$$

4 DADOS

Os dados que utilizamos neste trabalho provêm de três fontes distintas. Uma delas é a Pesquisa Nacional por Amostra de Domicílio (PNAD) do Instituto Brasileiro de Geografia e Estatística – IBGE realizadas em 1982, 1995 e 2001. A outra é o Sistema Nacional de Avaliação da Educação Básica (SAEB) realizado em 1995 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. A terceira fonte é o Censo Demográfico realizado em 2000 pelo IBGE³.

O IBGE a cada ano apresenta o resultado da PNAD para o conjunto do país, com informações básicas para o estudo e o planejamento do desenvolvimento sócio-econômico nacional, abrangendo características gerais da população, migração, educação, trabalho, rendimento e fecundidade, bem como famílias e domicílios. São informações para grande região, unidade da federação e região metropolitana.

O SAEB procura apoiar os municípios, os estados e a União na formulação de políticas que visam a melhoria da qualidade do ensino, coletando informações sobre alunos, professores e diretores de escolas públicas e privadas em todo o Brasil. Participam da

³ Sites do IBGE – www.ibge.gov.br – e do INEP – www.inep.gov.br .

avaliação alunos da 4^a e 8^a séries do ensino fundamental e da 3^a série do ensino médio, que fazem provas de Língua Portuguesa e de Matemática. Eles também respondem a um questionário sobre seus hábitos de estudo e suas características sócio-culturais. Os professores e os diretores participam respondendo a questionários que informam sobre perfil e prática docentes, mecanismos de gestão e infra-estrutura escolar.

Os Censos populacionais constituem a única fonte de informação sobre a situação de vida da população nos municípios e localidades. As realidades locais, rurais ou urbanas, dependem dos Censos para serem conhecidas e atualizadas. Eles produzem informações imprescindíveis para a definição de políticas públicas estaduais e municipais e para a tomada de decisões de investimento, sejam eles provenientes da iniciativa privada ou de qualquer nível de governo. Desde 1960, tem sido utilizada amostragem probabilística na coleta dos dados dos Censos Demográficos brasileiros, a exemplo do que ocorre em outros países, como nos Estados Unidos e no Canadá. É esse procedimento que permite, quando se realiza o Censo, a ampliação e o aprofundamento dos temas abordados para obtenção de informações mais completas sobre as condições de vida nos municípios e localidades. A operação do Censo 2000 abrangeu aproximadamente 42 (quarenta e dois milhões) de domicílios sendo visitados pelos recenseadores em 5507 (cinco mil, quinhentos e sete) municípios do país.

A fim de analisar a relação entre o desempenho escolar e os salários no Brasil formamos pseudo-painéis (bancos de dados) que acompanham a geração nascida em 1977 e 1978 em vários momentos de sua vida. Assim como em Dahl (2004), assumimos que os salários médios, as proficiências e as utilidades médias são os mesmos para tipos de pessoa similares. Utilizamos um vetor de variáveis contendo os três atributos considerados relevantes sobre um tipo de pessoa: sexo, cor e Unidade da Federação. Dessa forma, os indivíduos similares são agrupados por células, formando um modelo de pseudo-painel em que a geração é acompanhada ao longo do tempo.

Para formar os bancos de dados, através da PNAD de 2001 e do Censo 2000 obtemos informações sobre os indivíduos das coortes de 1977 e 1978, que possuíam renda, tinham pelo menos 10 anos de estudo e não migraram de 1995 em diante. A cada um desses bancos de dados é agrupada a nota média dessa geração no SAEB de 1995. São agrupadas ainda algumas variáveis características dessa geração que estão nas PNADs de 1982 e de 1995.

Todas as variáveis são agrupadas por células, a partir de suas médias, por três vetores características: sexo, cor e Unidade da Federação, gerando 108 observações (células) em cada um dos bancos de dados. Assim, definimos as variáveis: *SEXO* (homem e mulher), *RAÇA* (branca e outras) e *UNIDADES DA FEDERAÇÃO* para formar essas células representando os

108 tipos de indivíduos cuja geração acompanhamos em três etapas da vida. Escolhemos esses três vetores característica devido a resultados amplamente divulgados para o Brasil e outros países sobre as diferenças salariais entre homens e mulheres, entre raças [O'Neill (1990)], e entre unidades da federação, já que o retorno à educação não é equalizado entre os estados [Dahl (2002)]. Nas Figuras 1 e 2 detalhamos a composição da amostra, indicando os dados utilizados para a formação dos pseudo-painéis e suas características.

FIGURA 1 – FORMAÇÃO DOS PSEUDO-PAINEIS COM A PNAD DE 2001

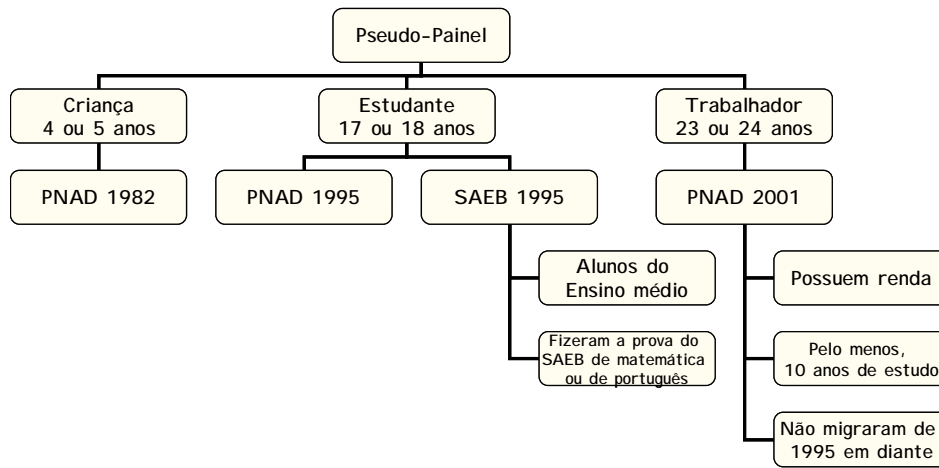
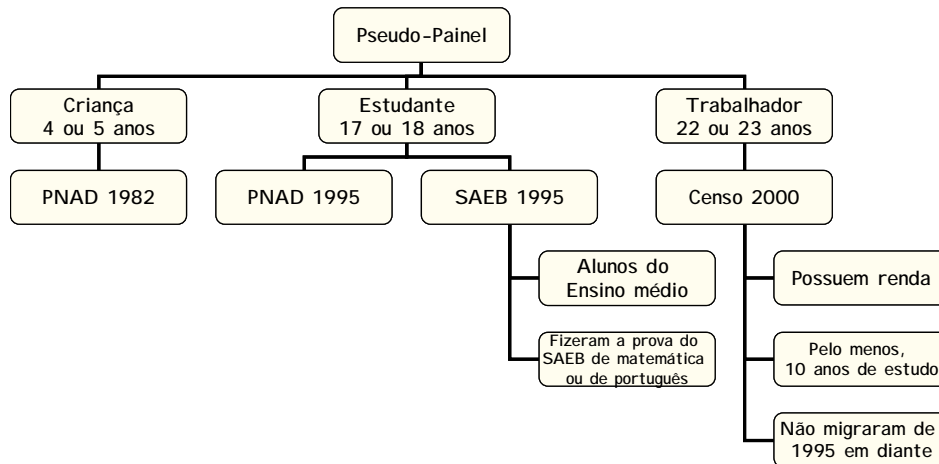


FIGURA 2 – FORMAÇÃO DOS PSEUDO-PAINEIS COM O CENSO DE 2000



A amostra utilizada da PNAD de 2001 e do Censo de 2000 representa os indivíduos na fase em que estão no mercado de trabalho. Esta abrange os jovens de 23 e 24 anos de idade (PNAD), e 22 e 23 anos de idade (Censo Demográfico), com 10 ou mais anos de estudo e que

não migraram após 1995, ano em que os jovens dessa geração fizeram as provas de proficiência do SAEB. Isto é, compõe a amostra aqueles que responderam que nasceram na Unidade da Federação onde vivem e trabalham, e aqueles que não nasceram nessa Unidade da Federação, mas nela moram sem interrupção desde 1995. Com relação aos salários, são usados nas regressões o logaritmo do salário real horário (*LWRH*), e nos gráficos, para facilitar a visualização, o logaritmo do salário real mensal. O deflator utilizado é o INPC para PNAD e para o Censo. Os salários também foram ponderados pelo custo de vida das regiões⁴.

O gráfico 1 apresenta, por Unidade da Federação, os salários médios (em logaritmo) recebidos em 2001 (dados da PNAD) pelos jovens brasileiros de 23 e 24 anos de idade, e pelos jovens dessas coortes com 10 ou mais anos de estudo e que não migraram após 1995 (amostra selecionada). Em todos os estados o salário médio dos jovens com 10 ou mais anos de estudo que não migraram após 1995 é maior que o salário médio dos jovens dessa geração. Isso ocorre dado o elevado nível educacional dos indivíduos que compõe a amostra selecionada. Comparando as regiões brasileiras, a região Nordeste destaca-se pela a maior diferença salarial entre as amostras, e a região sul pela a menor diferença.

Além disso, nesse primeiro gráfico os diferenciais de salários entre os estados são evidentes. Enquanto os salários reais mensais dos jovens pertencentes à amostra selecionada de São Paulo, Distrito Federal e Amapá eram, em média, superiores a 6.4 em logaritmo, no Piauí e no Ceará não passavam de 5.8. Destacamos também as discrepâncias regionais. As menores médias salariais de 2001 ocorreram nas regiões Norte e Nordeste, sendo que em todos os estados do Nordeste os salários médios em logaritmo para a amostra selecionada eram inferiores a 6.0, e para os jovens de 23 e 24 anos de idade como um todo, não passavam de 5.5, ambos abaixo das respectivas médias nacionais. Por outro lado, os salários médios na região Sul eram, em logaritmo, aproximadamente, 6.2 para a amostra selecionada e 6.0 quando consideramos todos os jovens das coortes analisadas, ambos superiores as respectivas médias nacionais. Entre os estados da região Sudeste também ocorrem desigualdades salariais significantes, enquanto São Paulo tinha uma das maiores médias salariais do Brasil em 2001, 6.4 para a amostra selecionada e 6.17 para os jovens em geral, no Espírito Santo essas médias salariais eram inferiores as médias nacionais, 5.9 e 5.65, respectivamente.

⁴ Menezes, T. e Azzoni, C. “Convergência de renda real e nominal entre as regiões metropolitanas brasileiras: uma análise de dados de painel”, XXVIII Encontro da Anpec, Campinas, 2000.

GRÁFICO 1 – SALÁRIO REAL MENSAL EM 2001 POR UF

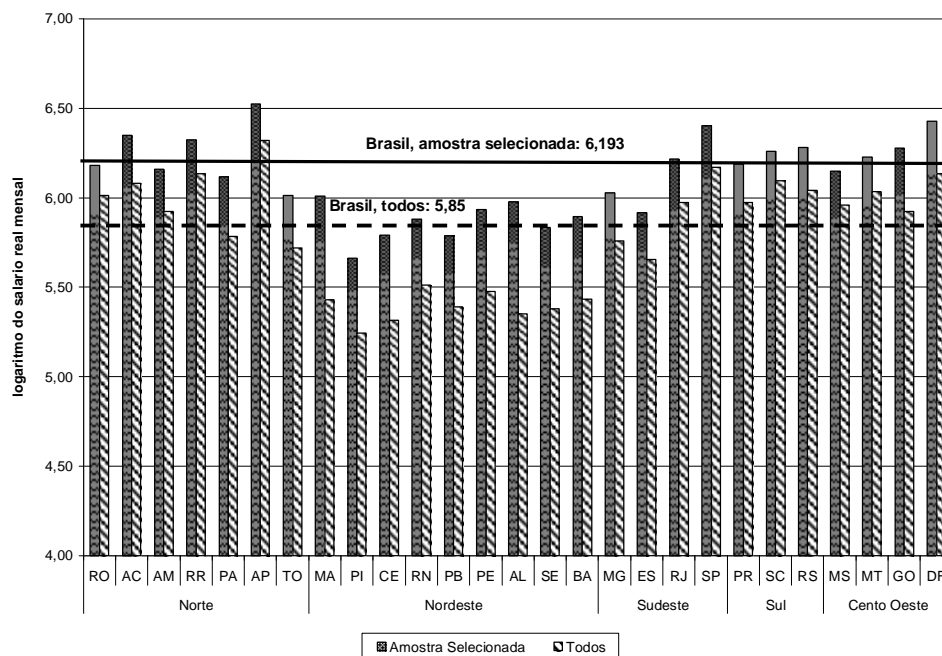
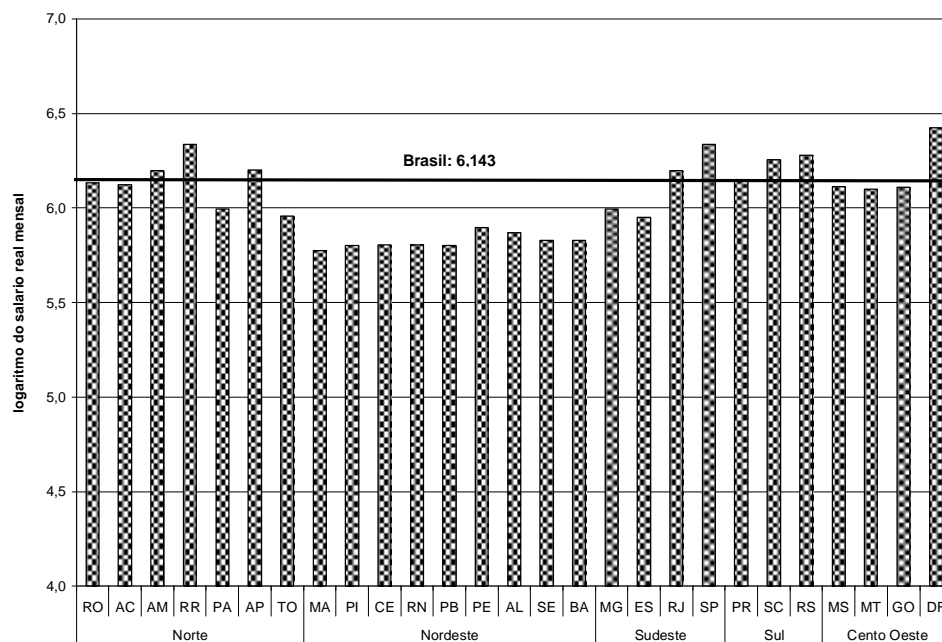


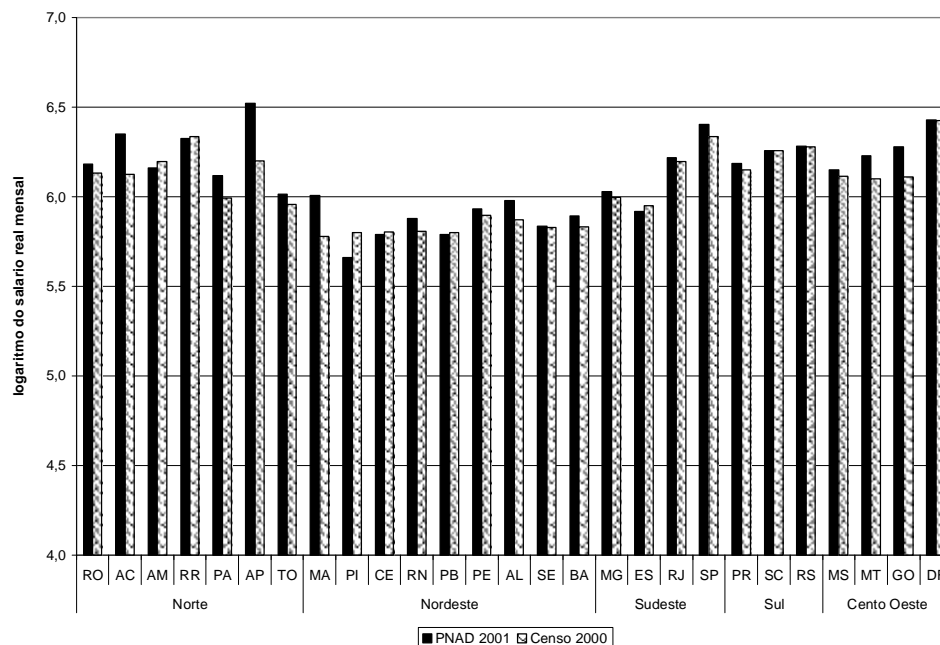
GRÁFICO 2 – SALÁRIO REAL MENSAL EM 2000 POR UF



O gráfico 2 destaca os salários em logaritmo recebidos em 2000 (dados do Censo Demográfico), em média, em cada uma das Unidades da Federação pelos jovens brasileiros de 22 e 23 anos, com 10 ou mais anos de estudo e que não migraram após 1995 (amostra selecionada). De acordo com os dados os salários mensais no Brasil em 2000 eram, em média,

equivalentes a 6.14 em logaritmo. Mesmo deflacionando os salários por índices regionais, as desigualdades salariais entre os estados e as regiões novamente são nítidas. Os salários da região Nordeste eram, em média, os menores do Brasil, em torno de 5.8 em logaritmo, enquanto na região Sul superavam os 6.2 em média. As Unidades da Federação com as maiores médias salariais em 2000 foram São Paulo, Distrito Federal e Roraima. Destacamos ainda o fato de que entre os estados do Nordeste as médias salariais são muito parecidas, enquanto na região Sudeste existe uma desigualdade salarial entre seus estados.

GRÁFICO 3 – SALÁRIO REAL MENSAL POR UF: PNAD VS CENSO



O gráfico 3 compara os salários médios em logaritmo recebidos, em média, em 2000 (dados do Censo Demográfico) e em 2001 (dados da PNAD) em cada uma das Unidades da Federação pelos jovens brasileiros que nasceram em 1977 e 1978, que têm 10 ou mais anos de estudo e que não migraram após 1995 (amostra selecionada). Em 21 dos 27 estados brasileiros houve expansão real dos salários. Comparando os salários em logaritmo, a maior retração salarial de 2000 para 2001 ocorreu no estado do Piauí, 2.46%. Nos demais estados em que os salários reais de 2001 foram inferiores aos de 2000, a retração em logaritmo foi inferior a 0.6%. Por outro lado, a maior expansão real dos salários nesse período, comparados em logaritmo, foi no Amapá, quase 5%, seguido pelo Maranhão e Acre, ambos em torno de 3,6%. Assim, comparando os salários reais de 2000 e de 2001, com dados do Censo e da PNAD,

respectivamente, ocorreu no Brasil uma expansão em logaritmo de 0.81% nos salários dos jovens não migrantes de 17 e 18 anos de idade, com 10 ou mais anos de estudo.

A partir dos dados da PNAD de 2001 e do Censo Demográfico de 2000 criamos duas variáveis binárias que identificam as características educacionais dos jovens das coortes de 1977 e 1978. Uma delas é uma variável de educação medida em anos de estudo, que identifica os indivíduos com 12 ou mais anos de estudo completos (*EDUCA_15*), e a outra é uma variável de frequência escolar que indica se esses jovens freqüentavam ou não escola nesses anos (*FREQ.ESCOLA*).

As duas variáveis cruciais do modelo que utilizamos para corrigir o viés de seleção provêm da PNAD de 1995 (correção do viés educacional), e da PNAD de 2001 e do Censo Demográfico de 2000 (correção do viés de migração). Assim como em Dahl (2002), usamos uma variável representando a porcentagem de indivíduos de 17 e 18 anos de idade que em 1995 tinham exatamente 10 anos de estudo e freqüentavam a escola (*VIES EDUCACIONAL*), isto é, a porcentagem de indivíduos dessa coorte aptos a fazer as provas de proficiência do SAEB naquele ano. A outra variável representa a porcentagem de pessoas das coortes de 1977 e 1978 que não migraram da Unidade da Federação em que viviam em 1995 (*VIES MIGRAÇÃO*).

Sendo a educação uma variável endógena, isto é, correlacionada com o termo aleatório, a introdução dessa variável ao modelo gera inconsistência do estimador de Mínimos Quadrados Ordinários do verdadeiro retorno adicional de salário relacionado à educação. A endogeneidade é causada por viés de variável omitida, erro de medida ou simultaneidade e viola a hipótese chave de consistência do estimador de MQO que é a condição de ortogonalidade populacional.

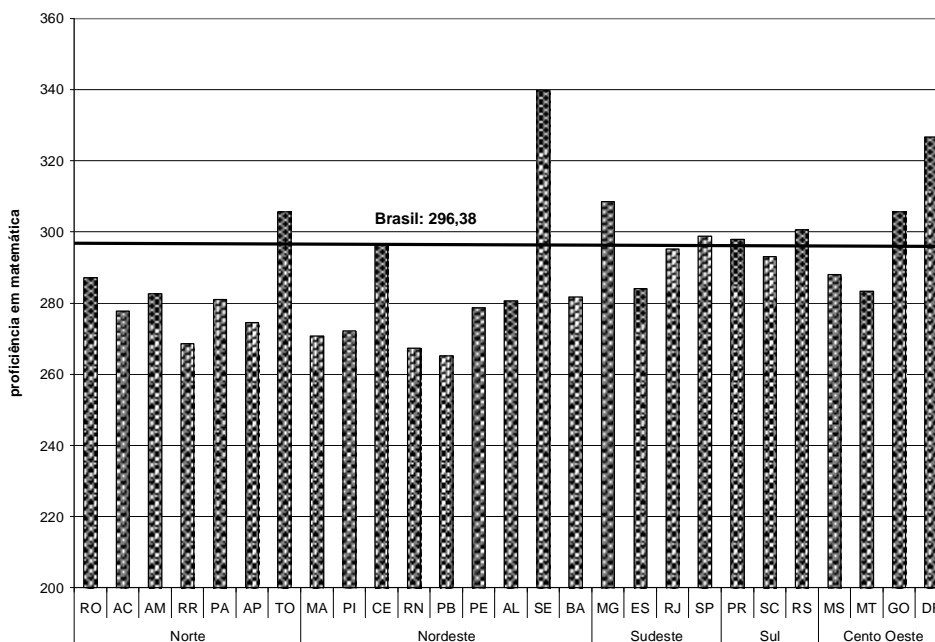
Em um modelo linear múltiplo, $y_i = b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + \dots + b_kx_k + u_i$, a hipótese de consistência é: $E(X'u) = 0 \Rightarrow \text{exogeneidade}$, que equivalente à $E(u) = 0, Cov(x_j, u) = 0, \forall j = 1, \dots, k$. Assim, no modelo desse trabalho, a variável *EDUCA_15* é endógena. Conseqüentemente:

$$E(u) = 0, Cov(x_j, u) = 0, Cov(x_k, u) \neq 0 \Rightarrow \text{endogeneidade} \Rightarrow \text{MQO inconsistente.}$$

O viés de variável omitida pode ser corrigido, ou pelo menos atenuado, utilizando uma variável *proxy*. A variável *proxy* que usamos nesse trabalho é a educação do pai quando os indivíduos eram crianças, tinham 4 e 5 anos de idade. Essa variável provêm da PNAD de 1982, e representa a educação média do chefe do domicílio em 1982 (*MEDCH_82*).

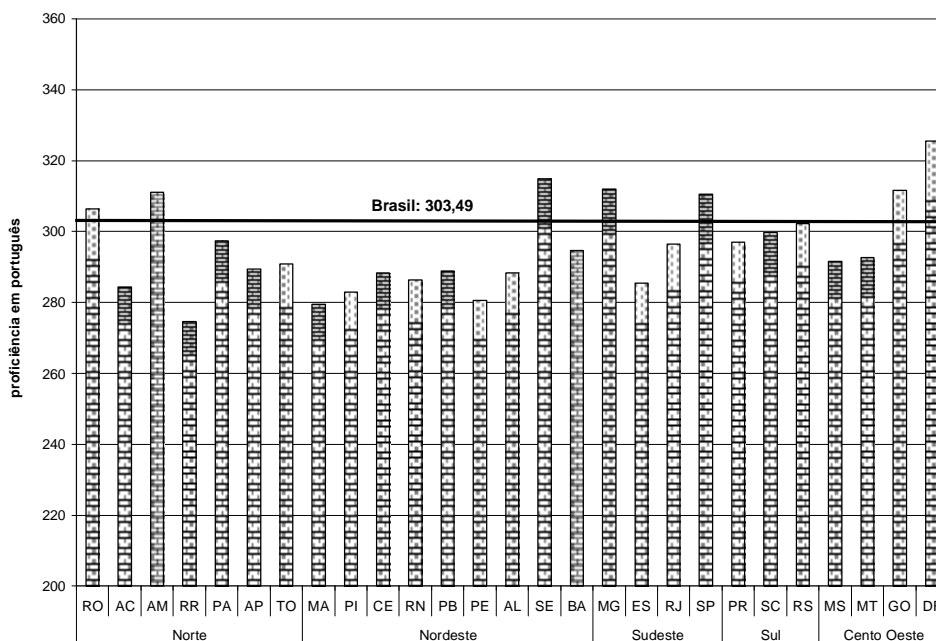
No gráfico 4 mostramos o desempenho médio em 1995 em cada estado brasileiro dos alunos de 17 e 18 anos no exame de proficiência em matemática realizado pelo INEP. As maiores pontuações ocorreram em Sergipe, Tocantins, Distrito Federal, Goiás e Rio Grande do Sul, todas elas, em média, superiores a 300 pontos. As menores notas, inferiores a 270 pontos, ocorreram em Roraima, Rio Grande do Norte e Pernambuco. Assim como acontece com os salários recebidos pelos jovens, o desempenho escolar médio dos alunos apresenta desigualdades regionais. Em todos os estados da região Sul os resultados médios foram superiores a 290 pontos, enquanto que nos estados das regiões Norte e Nordeste as notas médias foram inferiores a 280 pontos, com poucas exceções.

GRÁFICO 4 – PROFICIÊNCIA EM MATEMÁTICA POR UF EM 1995



No gráfico 5 mostramos o desempenho médio dos alunos das coortes de 1977 e 1978 em cada Unidade da Federação no exame de proficiência em língua portuguesa realizado pelo SAEB em 1995. Os estados que obtiveram o melhor desempenho nessa matéria foram Sergipe, Minas Gerais, Distrito Federal, Amazonas e Goiás, todos com notas superiores a 310 pontos em média. As médias inferiores a 280 pontos ocorreram em Roraima e no Maranhão. O desempenho escolar médio dos alunos em português também apresenta desigualdades regionais. Na região Sul todos os estados tiveram resultados médios superiores a 297 pontos, enquanto que nas regiões Norte e Nordeste as notas médias dos estados foram inferiores a 290 pontos, com poucas exceções.

GRÁFICO 5 – PROFICIÊNCIA EM PORTUGUÊS POR UF EM 1995



Os gráficos 4 e 5 apresentam o estado de Sergipe como tendo as maiores notas, em média, no exame de proficiência em matemática e língua portuguesa. Assim como os estados de Tocantins e Amazonas com notas altas em matemática e em português, respectivamente. Apesar de parecer estranho, isso pode ser explicado pelo fato de que nessas Unidades da Federação apenas uma pequena parcela da população (uma “elite” não representativa) atinge tal nível educacional (10 anos de estudo). Por outro lado, nos estados maiores, nos quais foram aplicadas políticas de progressão continuada como em São Paulo, uma grande massa populacional chega ao 3º. ano do ensino médio, tendo, em média, proficiência inferior.

Nos gráficos 6, 7, 8 e 9 cruzamos os gráficos de salários (gráficos 1 e 2) com os de proficiência (gráficos 4 e 5) por Unidade da Federação, mostrando a relação entre a nota de proficiência obtida pelos jovens aos 17 ou 18 anos de idade e os salários recebidos por eles no mercado de trabalho.

Os quatro gráficos mostram uma relação positiva para a geração de 1977 e 1978 entre as notas médias obtidas por esta nos exames de proficiência em matemática e em português em 1995 e os salários médios recebidos pela mesma cinco e seis anos depois, em 2000 e 2001. Os gráficos indicam que, em média, nas regiões em que os jovens que têm melhor desempenho escolar ao final do ensino médio, a remuneração ao trabalho é maior. Isto é, em média, as maiores notas nos exames de proficiência estão relacionadas aos maiores salários futuros.

GRÁFICO 6 – PROFICIÊNCIA EM MATEMÁTICA (1995) VS SALÁRIO (2001) POR UF

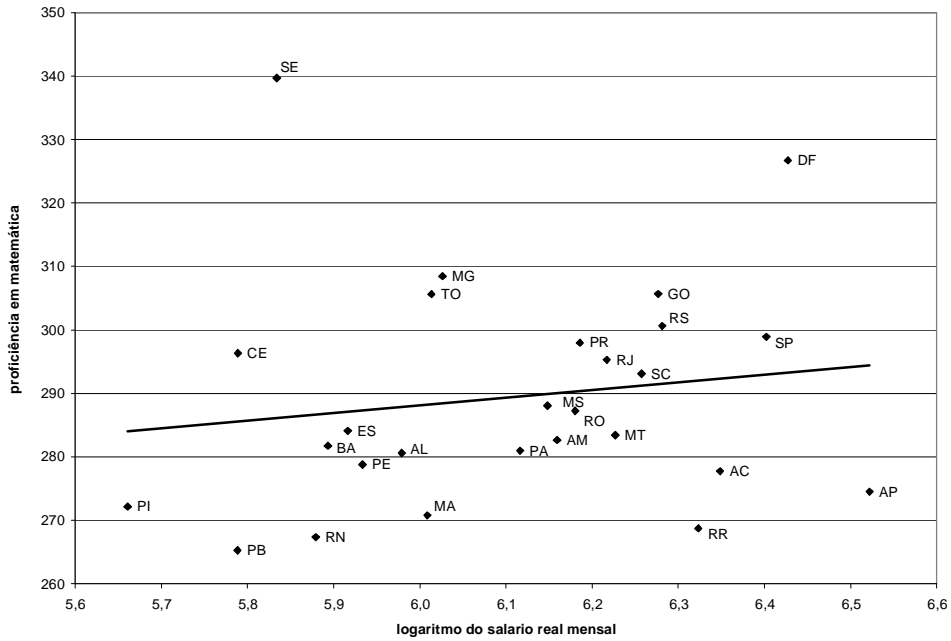


GRÁFICO 7 – PROFICIÊNCIA EM PORTUGUÊS (1995) VS SALÁRIO (2001) POR UF

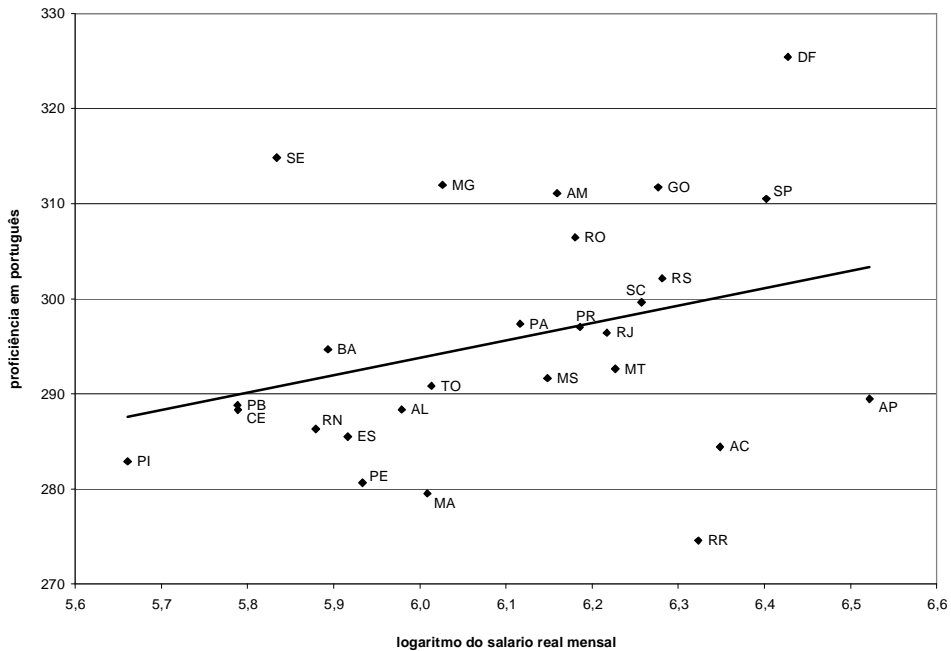


GRÁFICO 8 – PROFICIÊNCIA EM MATEMÁTICA (1995) VS SALÁRIO (2000) POR UF

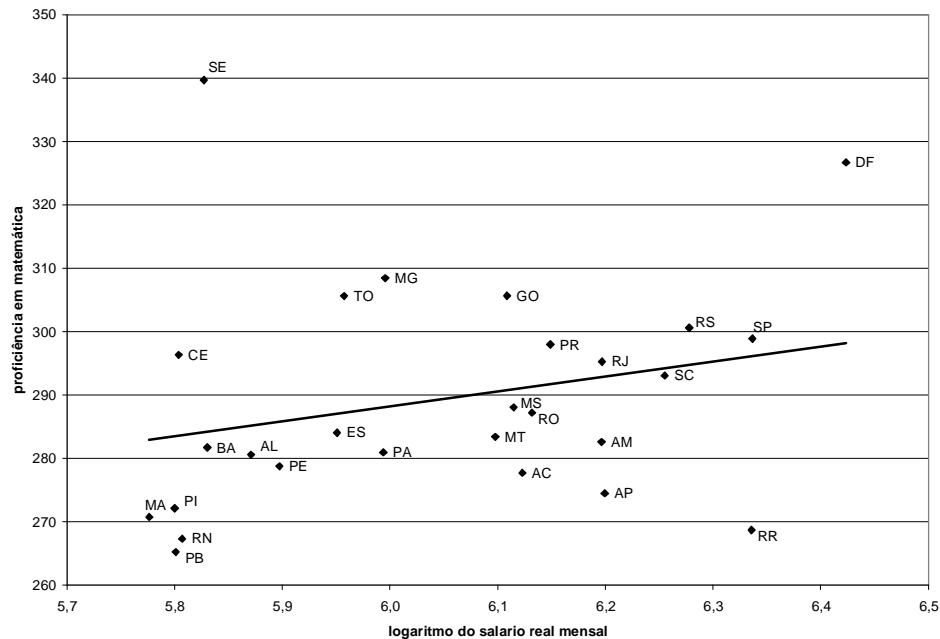
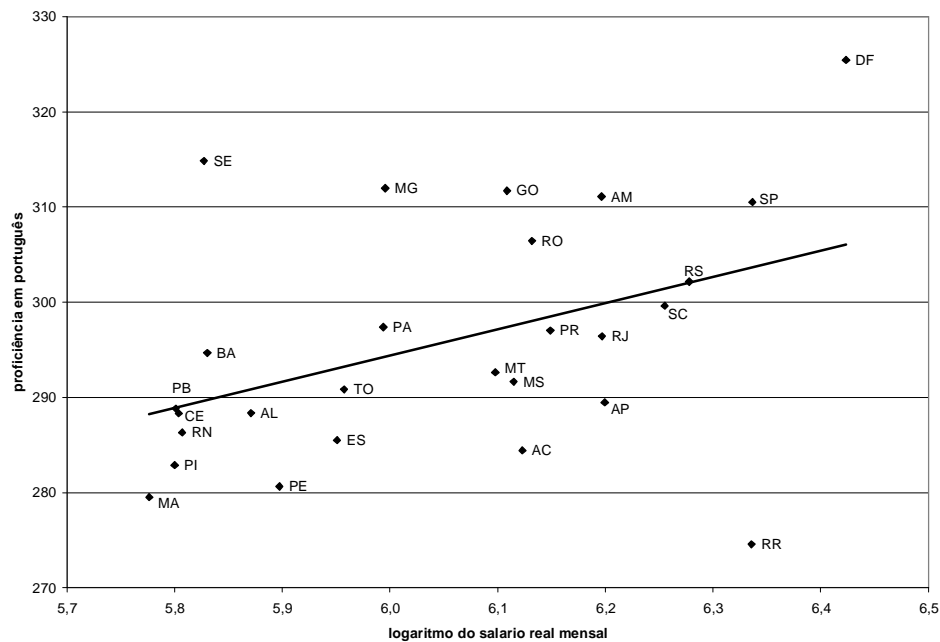


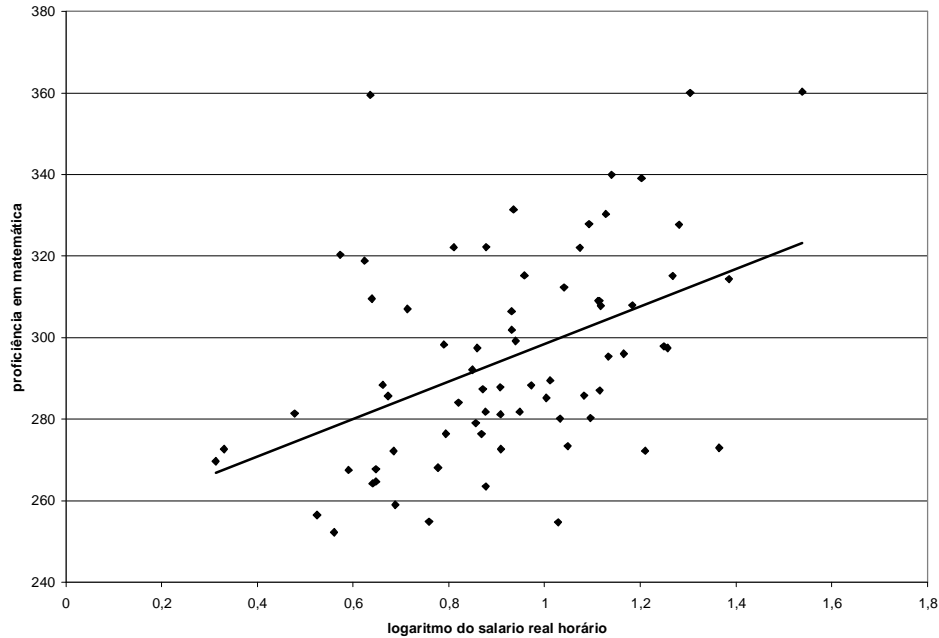
GRÁFICO 9 – PROFICIÊNCIA EM PORTUGUÊS (1995) VS SALÁRIO (2000) POR UF



Os gráficos 10, 11, 12 e 13, repetem o exercício dos quatro gráficos anteriores, mas por células. Cada célula representa um tipo de indivíduo de acordo com o sexo, a cor e a Unidade da Federação. São indivíduos que nasceram em 1977 e 1978 e que não migraram a partir de 1995, isto é, que fizeram a prova de proficiência em 1995 na mesma Unidade da

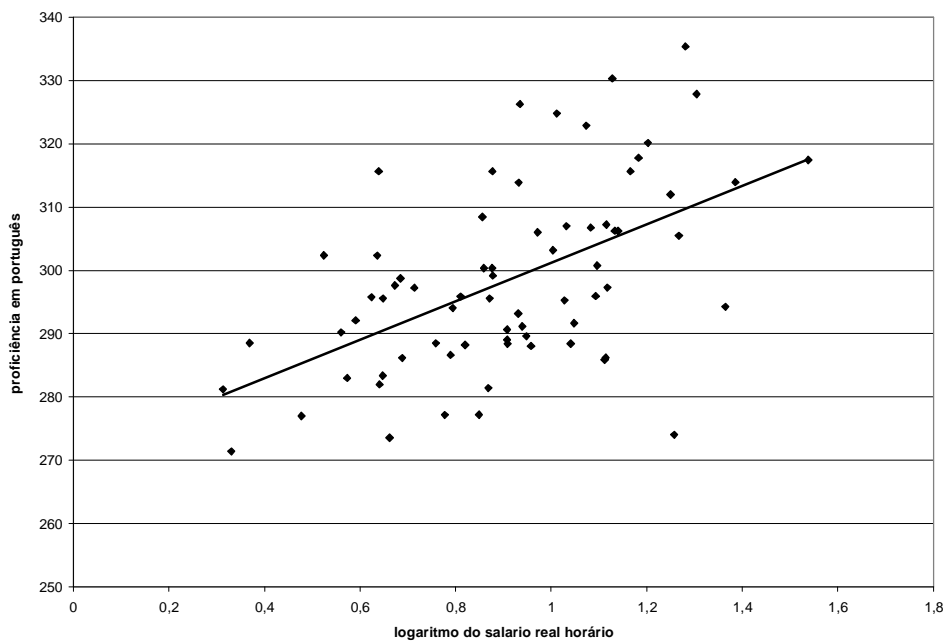
Federação que trabalham em 2000/2001. São consideradas apenas as células compostas por pelo menos 10 indivíduos, para garantir a representatividade das mesmas.

GRÁFICO 10 – PROFICIÊNCIA EM MATEMÁTICA (1995) VS SALÁRIO (2001) POR CÉLULA



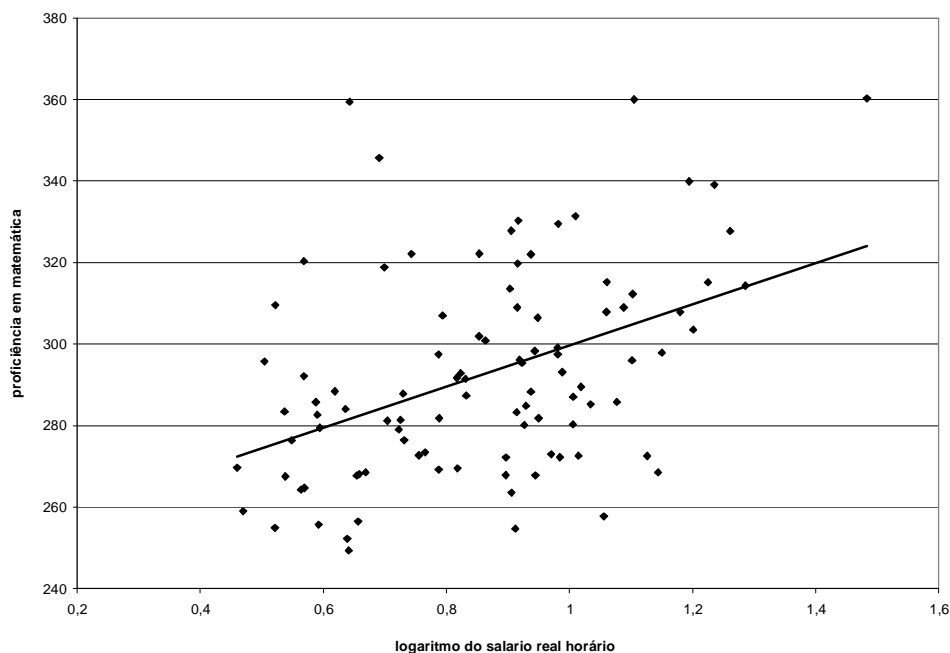
NOTA: o gráfico mostra as 72 células com 10 ou mais indivíduos.

GRÁFICO 11 – PROFICIÊNCIA EM PORTUGUÊS (1995) VS SALÁRIO (2001) POR CÉLULA



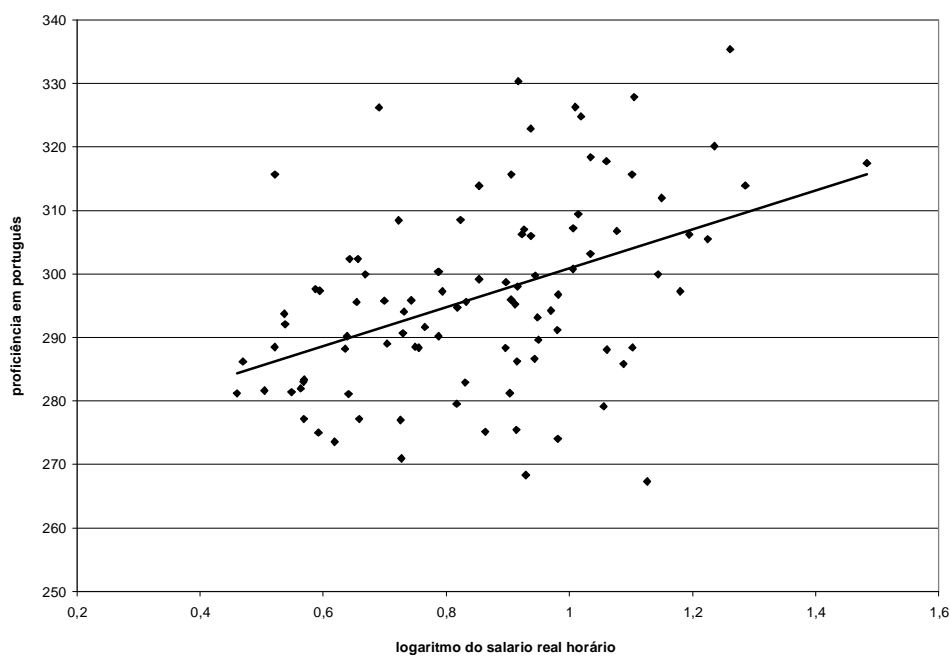
NOTA: o gráfico mostra as 72 células com 10 ou mais indivíduos.

GRÁFICO 12 – PROFICIÊNCIA EM MATEMÁTICA (1995) VS SALÁRIO (2000) POR CÉLULA



NOTA: o gráfico mostra as 99 células com 10 ou mais indivíduos.

GRÁFICO 13 – PROFICIÊNCIA EM PORTUGUÊS (1995) VS SALÁRIO (2000) POR CÉLULA



NOTA: o gráfico mostra as 97 células com 10 ou mais indivíduos.

Observamos que mesmo controlando por outras características a relação positiva entre o salário e o desempenho escolar se mantém. A significância dessas quatro relações apresentadas é verificada em seguida nas estimações do modelo de pseudo-panel.

Nesse trabalho utilizamos apenas as células da PNAD, do SAEB e do Censo Demográfico compostas por, pelo menos, 10 indivíduos, a fim de garantir maior representatividade dessas células. A tabela 1 mostra o número de células em que isso ocorreu em cada um dos bancos de dados que utilizamos. Os bancos de dados formados com a amostra do Censo Demográfico têm mais células que os bancos de dados formados com a amostra da PNAD, pois a primeira pesquisa abrange muito mais pessoas que a segunda. Assim, todas as 108 células formadas a partir do Censo têm pelo menos 50 indivíduos, enquanto algumas das células formadas com a amostra da PNAD apresentam menos de 10 indivíduos e são descartadas. Entretanto, o banco de dados com a amostra do Censo não ficou com as 108 células, pois com a amostra do SAEB de 1995 também são formadas células pouco representativas que são excluídas do modelo.

TABELA 1 – NÚMERO DE CÉLULAS COM 10 OU MAIS INDIVÍDUOS POR PSEUDO-PAINEL

SAEB 1995	PNAD 2001	Censo 2000
Português	72	97
Matemática	72	99
Português Resíduo	65	82
Matemática Resíduo	65	82

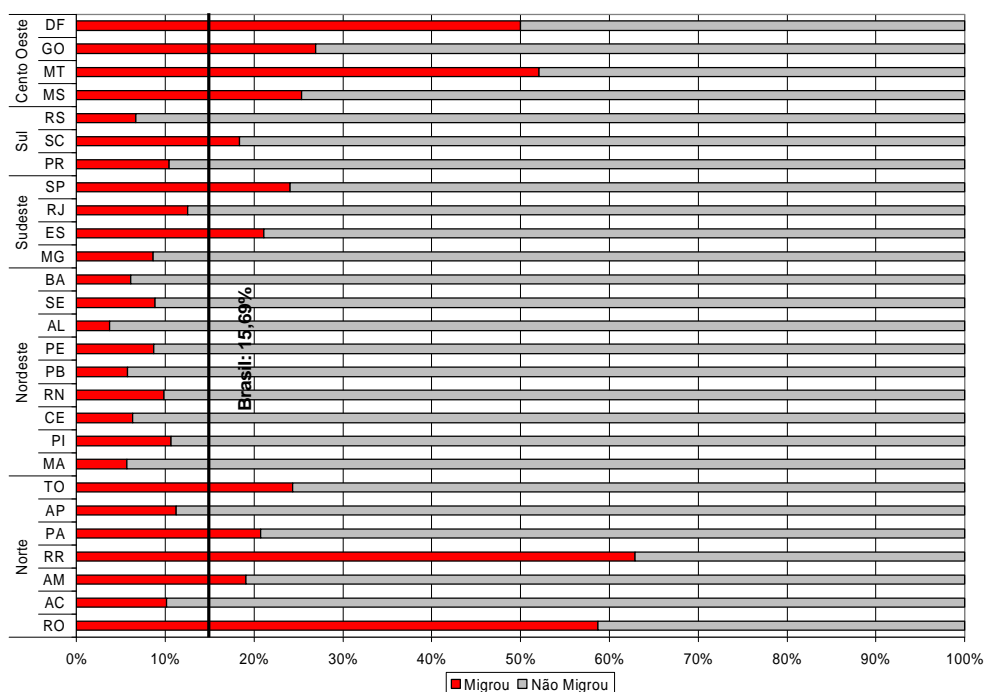
4.1 Análise da Migração no Brasil da Amostra Seleccionada

Utilizando a PNAD e o Censo Demográficos realizados pelo IBGE em 2001 e 2000, respectivamente, analisamos a migração dos jovens brasileiros que nasceram em 1977 e 1978. Examinamos a porcentagem de jovens que moram em cada estado brasileiro nesses anos, mas que não nasceram nesses estados, e a porcentagem de jovens que foram morar nesses estados depois de 1995, ano da realização da prova de proficiência do SAEB feita por essa geração aos 17 e 18 anos de idade. Essa segunda análise é exatamente a variável que utilizamos no modelo para corrigir o viés de migração da amostra selecionada que representa a porcentagem de não migrantes após 1995.

A partir dos dados da PNAD de 2001, o gráfico 14 mostra a porcentagem de pessoas que não nasceram na federação de residência. Observamos que em Rondônia, Roraima, Mato Grosso e Distrito Federal mais de 50% dos jovens das coortes de 1977 e 1978 não nasceram nos estados onde residem, valor bem acima da média nacional de migrantes, 15.7%. Já em

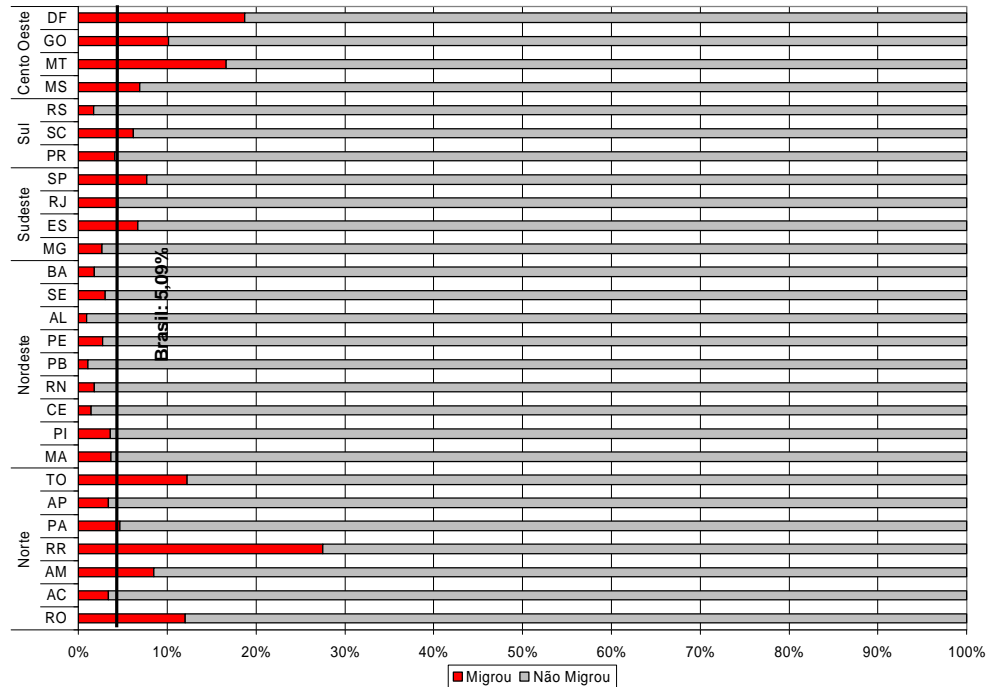
Alagoas menos de 3.8% dos jovens dessas coortes são proveniente de outros estados. Na região Centro Oeste, em todos os seus estados a média de migrantes é maior que a média do Brasil. Na região Nordeste, por sua vez, os estados tem as menores porcentagens de pessoas de 23 e 24 anos de idade que não nasceram no estado onde moram, menos de 10% em cada um. Isso pode ser explicado pelos baixos salários dos estados nordestinos que não atraem a população dos demais estados.

GRÁFICO 14 – PORCENTAGEM DE MIGRANTES POR UF (PNAD 2001)



O gráfico 15 mostra a porcentagem por Unidade da Federação de jovens de 23 e 24 anos de idade que migraram para o estado onde moram em 2001 após o ano de 1995. Em média, menos de 5.1% da população brasileira dessa geração migrou de 1995 em diante para a Federação de residência em 2001. Mais de 16% dos jovens residentes no Mato Grosso, Distrito Federal e Tocantins foram para os estados onde moram depois de 1995. Em contrapartida, em Alagoas menos de 1% das pessoas migraram para esse estado nordestino desse ano em diante. Os estados das regiões Centro Oeste e Norte são os que apresentam as maiores porcentagens de jovens migrantes após 1995 na sua população. Já os estados da região Nordeste têm as menores porcentagens de pessoas que para lá migraram de 1995 em diante. Em cada estado nordestino, em média, menos de 4% da população dessas coortes migraram depois de 1995 para o estado de residência em 2001.

GRÁFICO 15 – PORCENTAGEM DE MIGRANTES APÓS 1995 POR UF:
VIÉS DE MIGRAÇÃO (PNAD 2001)



Os gráficos 16 e 17 mostram a mesma análise dos dois gráficos anteriores usando a amostra do Censo de 2000. Segundo essa pesquisa, mais de 60% dos jovens das coortes de 1977 e 1978 que em 2000 moravam em Rondônia e em Roraima não nasceram nesses estados, assim como, cerca de 45% dos jovens dessa geração do Distrito Federal e do Mato Grosso. Essas médias são muito superiores a média nacional de migrantes que é de aproximadamente 15,7%. Por outro lado, no Rio Grande do Sul apenas 5% de seus jovens moradores não nasceram nesse estado sulista, assim como 6% dos que moram na Bahia e no Maranhão. Comparando as regiões, menos de 10% da população jovem nordestina não nasceu em seu estado de residência, enquanto nos estados do Centro Oeste do país mais de 23% dos jovens são migrantes nascidos em outras Federações.

De acordo com o gráfico 17, apenas os estados de Roraima e Distrito Federal têm mais de 20% de sua população composta por jovens de 22 e 23 anos de idade que para lá migraram de 1995 em diante. Já no Rio Grande do Sul, Bahia e Maranhão, esse número é inferior a 2% da população dessas coortes. A média nacional de jovens que migraram para o estado de residência em 2000 a partir de 1995 é 5,23%. Comparando as regiões, cada um dos estados nordestinos é formado por menos de 3,5% de jovens migrantes que lá mudaram após o ano de 1995. Na região Centro Oeste, por sua vez, mais de 6% dos residentes de seus estados são migrantes que neles chegaram após 1995.

GRÁFICO 16 – PORCENTAGEM DE MIGRANTES POR UF (CENSO 2000)

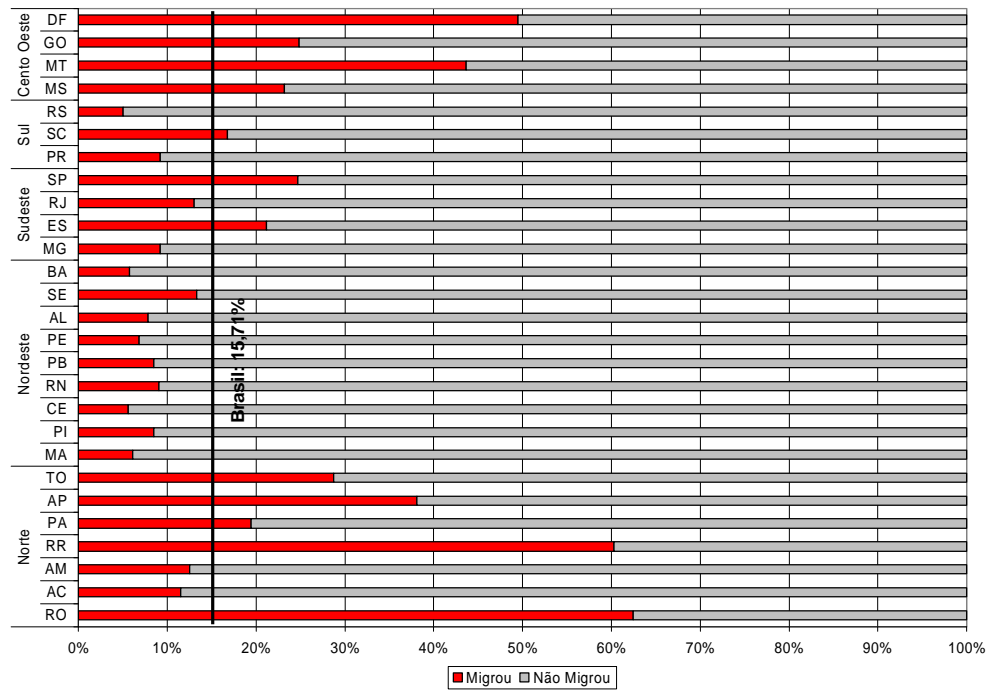
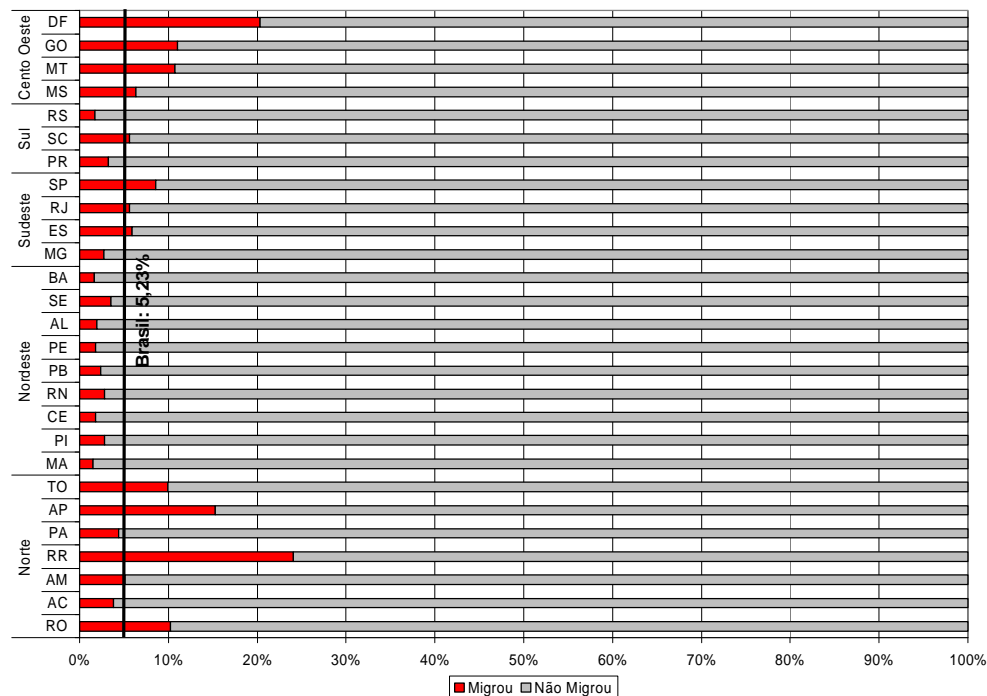


GRÁFICO 17 – PORCENTAGEM DE MIGRANTES APÓS 1995 POR UF: VIÉS DE MIGRAÇÃO (CENSO 2000)



As tabelas 2 e 3 mostram as origens e os destinos dos jovens brasileiros nascidos em 1977 e 1978 que migram entre os estados do país após o ano de 1995. De acordo essas tabelas, os estados da região Sudeste foram os destinos de mais de 48% das pessoas que

migraram de 1995 em diante do estado onde moravam, sendo que mais de 55% das pessoas migrantes que nasceram em algum estado dessa região migraram para outro da mesma. Mais de 55% dos nordestinos que deixaram o estado onde moravam em 1995 migraram para a região Sudeste. Os estados que menos atraíram migrantes foram os das regiões Norte e Sul. Para cada uma dessas regiões foram apenas 10% dos brasileiros que migraram de 1995 em diante.

TABELA 2 – DESTINO DOS JOVENS MIGRANTES APÓS 1995 (PNAD 2001)

em porcentagem Região de Origem	Região de Destino					Total
	Norte	Nordeste	Centro Oeste	Sudeste	Sul	
Norte	37,75	3,93	30,15	24,63	3,54	100
Nordeste	8,73	17,62	16,93	55,24	1,47	100
Centro Oeste	11,57	11,25	36,65	29,80	10,73	100
Sudeste	4,18	8,75	17,80	57,31	11,96	100
Sul	5,03	2,22	17,82	34,61	40,32	100
Total	9,58	12,08	19,52	48,76	10,06	100

TABELA 3 – DESTINO DOS JOVENS MIGRANTES APÓS 1995 (CENSO 2000)

em porcentagem Região de Origem	Região de Destino					Total
	Norte	Nordeste	Centro Oeste	Sudeste	Sul	
Norte	48,49	9,48	26,16	13,54	2,33	100
Nordeste	8,48	13,1	14,86	61,95	1,61	100
Centro Oeste	16,09	7,7	39,73	29,67	6,8	100
Sudeste	4,73	12,86	16,13	55,32	10,96	100
Sul	5,36	2,69	16,96	35,72	39,27	100
Total	10,42	11,2	17,7	52,27	8,42	100

4.2 Análise da Educação dos jovens no Brasil em 1995

Nesta seção analisamos o nível educacional dos jovens brasileiros de 17 e 18 anos de idade em 1995. Damos ênfase à representatividade nos estados dos jovens das coortes de 1977 e 1978 com pelo menos 10 anos de estudo que freqüentavam a escola, que são o foco desse trabalho.

Um indivíduo que começou a freqüentar a escola aos 6 ou 7 anos de idade e nunca repetiu deve ter 10 anos de estudo completos aos 17 ou 18 anos de idade. Entretanto, o gráfico 18 mostra que em 1995, das pessoas que nasceram em 1977 e 1978, 7% não têm sequer um ano de estudo completo, e que mais de 12,8% delas tem apenas o primeiro grau completo (4ª série do ensino básico). Isto é, em 1995 quase 35% da população brasileira com 17 e 18 anos de idade tinha menos de 5 anos de estudos completos quando deveriam estar terminando o ensino médio (cursando o 11º ano de estudo). Ainda segundo esse gráfico menos de 11,5%

dos jovens completaram o ginásio (8 anos de estudo), e menos de 10% dos jovens tem 10 anos de estudo completos, nível educacional que seria o correto se tivessem entrado na escola na idade certa e sido aprovados em todas as séries.

GRÁFICO 18 – NÍVEL EDUCACIONAL DOS JOVENS BRASILEIROS EM 1995

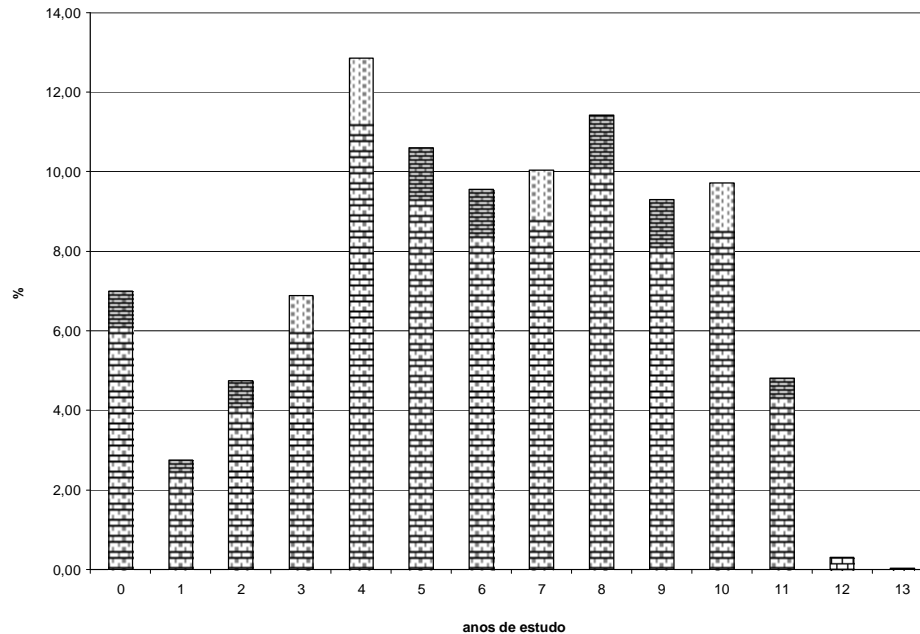
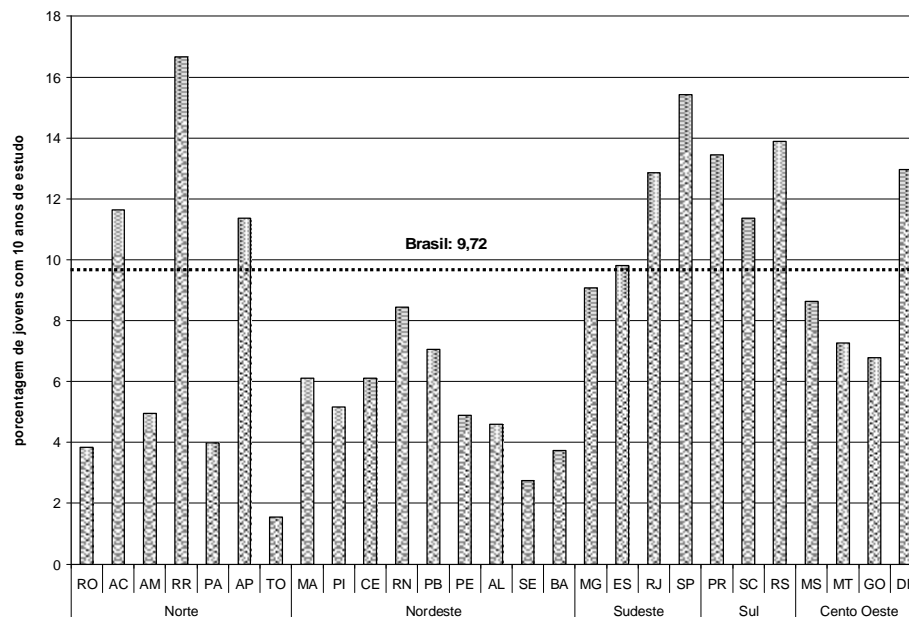


GRÁFICO 19 – PORCENTAGEM DE JOVENS POR UF COM 10 ANOS DE ESTUDO COMPLETOS

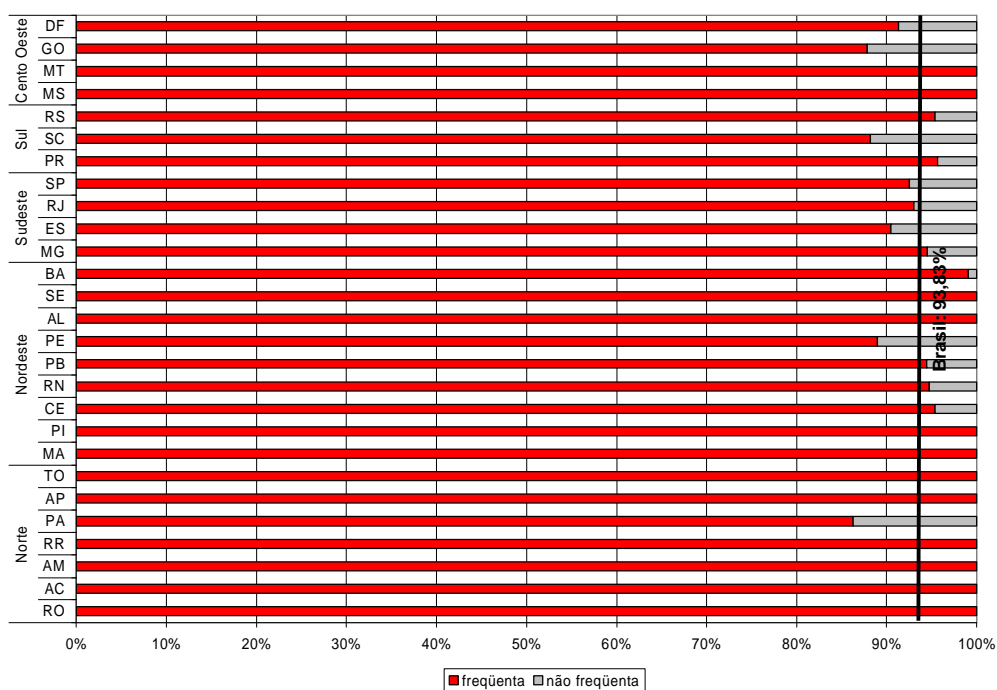


O gráfico 19 mostra a porcentagem em cada Unidade da Federação de jovens de 17 e 18 anos de idade com 10 anos de estudo completos em 1995. Os estados nordestinos, com exceção do Rio Grande do Norte, têm menos de 7% dos jovens das coortes de 1977 e 1978

com 10 anos de estudos. Já nos estados da região Sul a porcentagem de jovens com esse nível educacional é superior à média nacional que é de 93,83%. O mesmo ocorre com os estados da região Sudeste, com exceção de Minas Gerais onde apenas 9% dos jovens das coortes em análise têm 10 anos de estudo.

No gráfico 20 destacamos a frequência escolar desses jovens de 17 e 18 anos de idade com 10 anos de estudo em cada estado brasileiro. Esses jovens em 1995 deveriam estar frequentando o 11º ano de estudo que é a última série do ensino médio. A média nacional de frequência à escola desses jovens era de 93,83%. A região Norte se destacou pela elevada porcentagem de jovens na escola. Em seis dos sete estados dessa região todos os jovens com 10 anos de estudo frequentavam a escola. Na região Sudeste por outro lado, apenas o estado de Minas Gerais tem porcentagem de frequência à escola superior a média nacional.

GRÁFICO 20 – FREQUÊNCIA À ESCOLA DOS JOVENS EM 1995

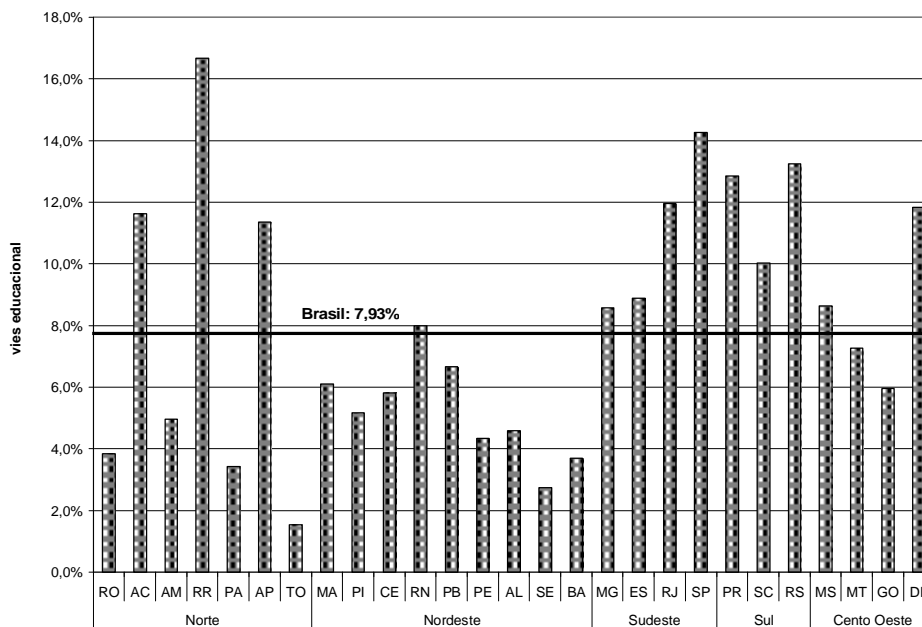


A amostra utilizada nesse trabalho é restrita aos indivíduos de 17 e 18 anos de idade que tinham exatamente 10 anos de estudo e frequentavam a escola em 1995. Isso porque esses indivíduos são aqueles aptos a fazer a prova de proficiência do SAEB, são jovens que nasceram em 1977 ou 1978 e que em 1995 frequentavam a última série do ensino médio. Como a amostra desse trabalho é selecionada e não aleatória, o modelo apresenta um viés de

seleção amostral que corrigimos utilizando uma variável que indica a porcentagem de jovens com essas características em cada célula.

O gráfico 21 apresenta a porcentagem de jovens na geração nascida em 1977 e 1978 com exatamente 10 anos de estudo que freqüentam a escola em 1995 por Unidade da Federação, que é a variável utilizada para corrigir o viés de seleção educacional do modelo. Isto é, representa a porcentagem de jovens em cada estado brasileiro com possibilidade de fazer a prova de proficiência dentre os jovens com 17 e 18 anos de idade. No Brasil cerca de 8% da população jovem tem essas características. Os estados de Roraima e São Paulo tinham mais de 14% da população aptas a fazer a prova do SAEB de 1995. Já o estado de Tocantins em 1995 era o único com menos de 2% da população jovem com exatamente 10 anos de estudo freqüentando a escola, e o estado de Sergipe com menos de 3%. Isso corrobora a análise que fizemos anteriormente de que esses estados obtêm elevadas médias nos exames de proficiência, pois apenas uma pequena parcela da população atinge tal nível educacional, enquanto no estado de São Paulo, em que foi aplicada a política de progressão continuada, uma grande parcela da população chega ao 3º. ano do ensino médio.

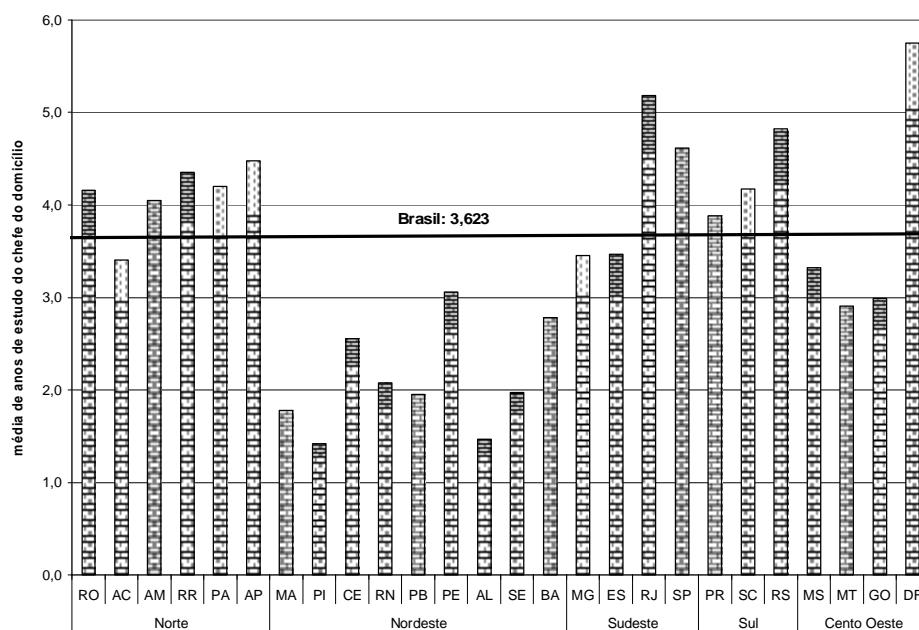
GRÁFICO 21 – VIÉS EDUCACIONAL



4.3 Análise da Educação dos pais da geração de 1977 e 1978 em 1982

Essa seção analisa a educação em 1982 do pai, chefe de domicílio, dos indivíduos da geração de 1977 e 1978, quando eles tinham 4 e 5 anos de idade. A educação do pai é uma característica importante para explicar o nível educacional do filho. Em geral, filhos de pais mais educados, com mais anos de estudo completos, tendem a ir mais longe nos estudos.

GRÁFICO 22 – EDUCAÇÃO DO CHEFE DO DOMICÍLIO EM 1982 POR UF



O gráfico 22 mostra que nos estados da região Nordeste, com uma exceção, os chefes de domicílio tinham, em média, menos de 2,8 anos de estudo em 1982. Ao contrário, nos estados da região Norte, com uma exceção, os chefes de domicílio tinham mais de 4 anos de estudo em média, valor superior à média nacional de 3,6 anos de estudo completos. Os estados do Distrito Federal e do Rio de Janeiro aparecem como os estados onde, em média, os chefes de domicílio têm maior escolaridade, 5,7 e 5,2 anos de estudo, respectivamente. Por outro lado, Piauí e Alagoas se destacam pelo baixo nível educacional dos chefes de domicílio, menos de 1,5 anos de estudo, em média. No Nordeste todos os estados apresentam média educacional dos chefes de domicílio inferior à média brasileira, ao contrário, no Sul todos os estados têm, em média, chefes com nível educacional superior à média nacional, o que também ocorre nos estados nortistas, com uma única exceção.

4.4 Sumário Estatístico

Nas tabelas 4 e 5 apresentamos as descrições, os número de observações, as médias e os desvios padrões, das variáveis utilizadas nesse trabalho. Na tabela 4 expomos as características das variáveis que provêm do SAEB de 1995 e que são utilizadas no primeiro estágio do modelo.

A partir das informações obtidas pelo SAEB de 1995 são obtidas as informações sobre as escolas, os professores, os diretores e os alunos. A amostra é restrita aos alunos das coortes analisados. As escolas são caracterizadas pelas variáveis: *REDE*, pública ou privada, *ZONA*, rural ou urbana, e as que indicam se escolas possuem ou não *COMPUTADORES*, laboratório de ciências (*LAB.CIÊNCIAS*) e/ou *BIBLIOTECA*. Sobre os professores e os diretores julgamos relevantes as informações sobre: a educação, o tempo no magistério e os salários desses profissionais. Com relação ao nível educacional, os professores e diretores são divididos em três grupos: aqueles com nível superior (*EDUC.PROF.1* e *EDUC.DIRETOR1*), os com pós-graduação (*EDUC.PROF.2* e *EDUC.DIRETOR2*), e os professores e diretores com níveis educacionais inferiores, usado como grupo de comparação. Pelo tempo no magistério, são separados da seguinte forma: professores e diretores com menos de 1 ano de experiência, o grupo de comparação, profissionais que estão de 1 a 10 anos no magistério (*PROF.MAGIST1* e *DIRETOR MAGIST1*), e aqueles há mais de 10 anos trabalhando no magistério (*PROF.MAGIST2* e *DIRETOR MAGIST2*). Com relação aos salários desses profissionais, os dados do SAEB de 1995 apresentam o salário como variável contínua. São geradas, a partir dessas, as variáveis discretas. Os intervalos salariais são: valores de 308 a 1023 (*SALÁRIO PROF1* e *SALÁRIO DIRETOR1*), salários entre 1024 e 2046 (*SALÁRIO PROF2* e *SALÁRIO DIRETOR2*), e valores acima de 2046 (*SALÁRIO PROF3* e *SALÁRIO DIRETOR3*). O grupo de comparação é o formado por profissionais cujos salários são inferiores a 308. Por fim, sobre os alunos são utilizados os dados sobre o logaritmo da proficiência em matemática ou em português (*PROFICIÊNCIA*) e o nível educacional das mães (*EDUCA_MÃE*), separadas entre aquelas que têm pelo menos o colegial completo e as que têm nível escolar inferior.

Na tabela 5 mostramos as variáveis que são utilizadas para estimar os modelos de pseudo-painel, em que é analisada a relação entre o desempenho escolar e os salários dos jovens brasileiros através do acompanhamento das gerações.

TABELA 4 – DESCRIÇÃO DAS VARIÁVEIS (SAEB)

SAEB 1995 Matemática				SAEB 1995 Português			
	Obs	Média	Desvio-Padrão		Obs	Média	Desvio-Padrão
HOMEM	2909	0,390	0,488	HOMEM	5803	0,366	0,482
BRANCO	2909	0,565	0,496	BRANCO	5803	0,560	0,496
PROFICIENCIA (em log)	2909	5,665	0,181	PROFICIENCIA (em log)	5803	5,685	0,186
IDADE	2909	17,407	0,491	IDADE	5803	17,427	0,495
MÃE COM COLEGIAL	2909	0,386	0,487	MÃE COM COLEGIAL	5803	0,368	0,482
SALARIO PROF1	2909	0,233	0,423	SALARIO PROF1	5197	0,234	0,424
SALARIO PROF2	2909	0,579	0,494	SALARIO PROF2	5197	0,576	0,494
SALARIO PROF3	2909	0,147	0,354	SALARIO PROF3	5197	0,148	0,355
SALARIO PROF4	2909	0,041	0,199	SALARIO PROF4	5197	0,042	0,200
PROF. MAGIST.1	2909	0,010	0,101	PROF. MAGIST.1	5803	0,012	0,109
PROF. MAGIST.2	2909	0,415	0,493	PROF. MAGIST.2	5803	0,422	0,494
PROF. MAGIST.3	2909	0,575	0,494	PROF. MAGIST.3	5803	0,565	0,496
EDUC.PROF.1	2909	0,061	0,240	EDUC.PROF.1	5154	0,066	0,248
EDUC.PROF.2	2909	0,667	0,471	EDUC.PROF.2	5154	0,658	0,474
EDUC.PROF.3	2909	0,272	0,445	EDUC.PROF.3	5154	0,276	0,447
SALARIO DIRETOR1	2909	0,042	0,200	SALARIO DIRETOR1	5345	0,049	0,216
SALARIO DIRETOR2	2909	0,366	0,482	SALARIO DIRETOR2	5345	0,369	0,483
SALARIO DIRETOR3	2909	0,396	0,489	SALARIO DIRETOR3	5345	0,389	0,488
SALARIO DIRETOR4	2909	0,197	0,398	SALARIO DIRETOR4	5345	0,193	0,395
DIRETOR MAGIST.1	2909	0,007	0,083	DIRETOR MAGIST.1	5803	0,005	0,068
DIRETOR MAGIST.2	2909	0,125	0,331	DIRETOR MAGIST.2	5803	0,124	0,329
DIRETOR MAGIST.3	2909	0,868	0,338	DIRETOR MAGIST.3	5803	0,871	0,335
EDUC.DIRETOR1	2679	0,050	0,217	EDUC.DIRETOR1	5747	0,058	0,234
EDUC.DIRETOR2	2679	0,622	0,485	EDUC.DIRETOR2	5747	0,620	0,485
EDUC.DIRETOR3	2679	0,328	0,470	EDUC.DIRETOR3	5747	0,322	0,467
ZONA URBANA	2909	0,986	0,119	ZONA URBANA	5803	0,986	0,119
REDE PRIVADA	2909	0,284	0,451	REDE PRIVADA	5803	0,285	0,451
BIBLIOTECA	2796	0,888	0,315	BIBLIOTECA	5689	0,883	0,321
LAB. CIÊNCIAS	2745	0,612	0,487	LAB. CIÊNCIAS	5634	0,609	0,488
COMPUTADORES	2762	0,280	0,449	COMPUTADORES	5695	0,282	0,450

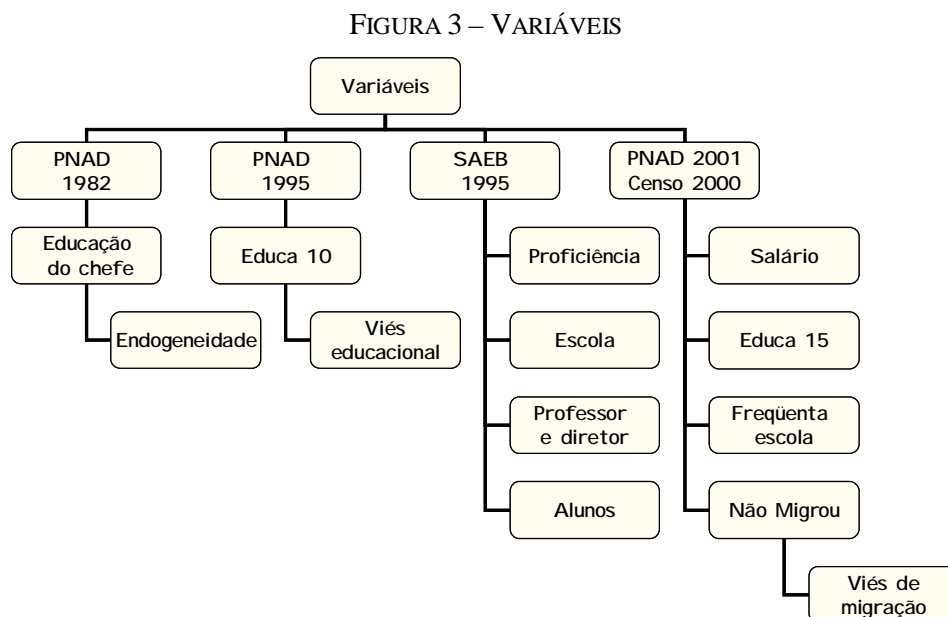
TABELA 5 – DESCRIÇÃO DAS VARIÁVEIS (PNAD, CENSO E SAEB)

PNAD + SAEB (PNAD's 1982, 1995, 2001+ SAEB 1995)												
Variáveis	Matemática			Português			Matemática Resíduo			Português Resíduo		
	Obs	Média	Desvio-Padrão	Obs	Média	Desvio-Padrão	Obs	Média	Desvio-Padrão	Obs	Média	Desvio-Padrão
HOMEM	72	0,472	0,5027	72	0,458	0,5018	65	0,462	0,5024	65	0,431	0,4990
BRANCO	72	0,458	0,5018	72	0,458	0,5018	65	0,477	0,5034	65	0,477	0,5034
LWRH	72	0,924	0,2521	72	0,912	0,2581	65	0,934	0,2518	65	0,904	0,2589
FREQ.ESCOLA	72	0,238	0,1072	72	0,242	0,1113	65	0,244	0,1043	65	0,251	0,1098
EDUCA (em anos de estudo)	72	11,553	0,4825	72	11,558	0,4874	65	11,588	0,4780	65	11,604	0,4836
EDUCA_15	72	0,227	0,1399	72	0,231	0,1466	65	0,238	0,1369	65	0,245	0,1447
PROFICIENCIA Matemática (em log)	72	5,669	0,0856				65	5,665	0,0842			
PROFICIENCIA Matemática	72	294,764	25,7170				65	293,647	25,1336			
PROFICIENCIA Português (em log)				72	5,681	0,0525				65	5,677	0,0513
PROFICIENCIA Português				72	298,017	14,9953				65	296,853	14,5456
NÃO_MIGRANTES (viés de migração)	72	0,936	0,0562	72	0,937	0,0563	65	0,934	0,0583	65	0,937	0,0585
EDUCA_10 (1995) (viés educacional)	72	0,078	0,0482	72	0,079	0,0481	65	0,083	0,0481	65	0,083	0,0481
MEDCH_82	70	3,308	1,3090	70	3,266	1,3015	63	3,342	1,3263	63	3,301	1,3233
RESIDUOS Matemática A+P+D+E							65	-0,005	0,0514			
RESIDUOS Português A+P+D+E										65	-0,0002	0,0429

PNAD + Censo + SAEB (PNAD's 1982, 1995 + Censo 2000 + SAEB 1995)												
Variáveis	Matemática			Português			Matemática Resíduo			Português Resíduo		
	Obs	Média	Desvio-Padrão	Obs	Média	Desvio-Padrão	Obs	Média	Desvio-Padrão	Obs	Média	Desvio-Padrão
HOMEM	99	0,475	0,5019	97	0,454	0,5004	82	0,451	0,5007	82	0,427	0,4977
BRANCO	99	0,465	0,5013	97	0,474	0,5019	82	0,476	0,5025	82	0,476	0,5025
LWRH	99	0,874	0,2108	97	0,868	0,2093	82	0,877	0,2145	82	0,870	0,2149
FREQ.ESCOLA	99	0,446	0,1734	97	0,444	0,1699	82	0,453	0,1789	82	0,452	0,1784
EDUCA (em anos de estudo)	99	11,349	0,2940	97	11,359	0,2931	82	11,367	0,3027	82	11,377	0,3072
EDUCA_15	99	0,186	0,0866	97	0,188	0,0872	82	0,190	0,0889	82	0,193	0,0905
PROFICIENCIA Matemática (em log)	99	5,661	0,0842				82	5,657	0,0818			
PROFICIENCIA Matemática	99	292,374	25,1578				82	291,062	24,3684			
PROFICIENCIA Português (em log)				97	5,675	0,0534				82	5,671	0,0509
PROFICIENCIA Português				97	296,166	15,2314				82	294,914	14,4345
NÃO_MIGRANTES (viés de migração)	99	0,938	0,0521	97	0,938	0,0527	82	0,936	0,0551	82	0,938	0,0544
EDUCA_10 (1995) (viés educacional)	99	0,076	0,0511	97	0,078	0,0502	82	0,081	0,0520	82	0,083	0,0496
MEDCH_82	95	3,226	1,2850	93	3,191	1,2602	78	3,276	1,3024	78	3,242	1,2764
RESIDUOS Matemática A+P+D+E							82	-0,008	0,0503			
RESIDUOS Português A+P+D+E										82	-0,003	0,0434

A variável de educação do pai (*MEDCH_82*) apresenta menos observações que as demais variáveis, pois em 1982 os dados do atual estado de Tocantins eram computados juntamente com os dados do estado de Goiás.

A figura 3 sintetiza as variáveis utilizadas nesse trabalho na formação dos pseudo-painéis. O organograma indica a origem de cada variável e o motivo de sua utilização no modelo.



5 RESULTADOS ECONOMÉTRICOS

Nas tabelas 6 a 9 reproduzimos os resultados das estimações da equação (1) do modelo de pseudo-painel⁵. A tabela 6 apresenta os resultados da estimação, considerando os jovens das coortes de 1977 e 1978 que fizeram a prova de proficiência em matemática do SAEB em 1995 e estão no mercado de trabalho em 2001 (dados da PNAD).

De acordo com as três primeiras colunas dessa tabela o salário dos jovens dessa geração aos 23 e 24 anos de idade é determinado pelo desempenho no exame de proficiência em matemática da mesma ao final do ensino médio aos 17 e 18 anos de idade. Da coluna (IV) em diante, em que são inseridas as variáveis de região, a proficiência não é significativa na determinação dos salários futuros na ausência da variável de educação medida em anos de

⁵ Os resultados apresentados em negrito nessas tabelas indicam as variáveis estatisticamente significantes a 10% na determinação da variável dependente.

estudo, mas seu impacto é positivo e significativo no salário futuro quando controlamos o modelo pelos anos de estudo (colunas VII e IX).

TABELA 6 – PROFICIÊNCIA EM MATEMÁTICA (1995) E SALÁRIOS (2001_PNAD)

	Variável Dependente: Logaritmo do Salário Real Horário								
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
PROFICIÊNCIA EM MATEMÁTICA	1,33 0,36	1,20 0,30	0,46 0,25	0,31 0,58	0,50 0,42	0,36 0,33	0,59 0,31	0,31 0,33	0,57 0,31
VIES MIGRAÇÃO		-1,45 0,42	-1,24 0,37	-0,24 0,85	-0,91 0,76	-0,83 0,65	-1,09 0,52	-211,61 186,48	-226,29 153,21
VIES MIGRAÇÃO ²								2,43 2,13	2,57 1,77
VIES MIGRAÇÃO ³								-0,09 0,08	-0,10 0,07
VIES EDUCACIONAL		1,55 0,45	-0,47 0,55	-0,30 0,74	-0,37 0,69	0,20 0,74	0,24 0,63	4,42 4,84	5,14 4,02
VIES EDUCACIONAL ²								-0,51 0,54	-0,58 0,44
VIES EDUCACIONAL ³								0,18 0,18	0,19 0,15
EDUCA_15			0,72 0,18		1,16 0,31		0,84 0,21		0,86 0,20
FREQ. ESCOLA			-0,38 0,24	-0,17 0,27	-0,59 0,22	-0,13 0,23	-0,54 0,21	-0,07 0,23	-0,49 0,22
MEDCH_82 (em log)			0,27 0,06	-0,13 0,26	0,02 0,23	0,25 0,09	0,25 0,08	0,24 0,10	0,24 0,08
HOMEM			0,13 0,05	0,11 0,08	0,17 0,05	0,11 0,06	0,15 0,05	0,13 0,05	0,16 0,05
BRANCA			0,06 0,06	0,30 0,10	0,09 0,12	0,14 0,05	0,02 0,06	0,13 0,06	0,01 0,06
NORDESTE						-0,08 0,08	-0,10 0,07	-0,05 0,09	-0,09 0,07
CENTRO OESTE						0,04 0,08	-0,08 0,08	0,07 0,08	-0,06 0,09
SUDESTE						-0,11 0,08	-0,19 0,07	-0,09 0,08	-0,18 0,08
SUL						-0,01 0,09	-0,08 0,08	0,02 0,10	-0,06 0,09
UF				SIM	SIM				
CONSTANTE	-6,61 2,06	-4,63 1,80	-0,91 1,56	-0,36 3,34	-1,15 2,49	-0,63 2,04	-1,71 1,82	60,02 53,97	63,60 43,96
R ²	0,204	0,418	0,710	0,756	0,848	0,683	0,747	0,697	0,762
No. OBSERVAÇÕES	72	72	70	70	70	70	70	70	70

Na coluna (II) as duas variáveis de correção do viés do modelo, de migração e educacional, são significantes e se relacionam negativamente e positivamente aos salários, respectivamente. Na coluna (III) são inseridas as variáveis que indicam a influência dos anos de estudo nos salários, da frequência escolar, da educação dos pais e de características

individuais como sexo e cor. De acordo com os resultados, os indivíduos que aos 23 e 24 anos de idade têm 12 ou mais anos de estudo recebem, em média, salários maiores que aqueles com nível educacional inferior. Além disso, a média educacional do chefe do domicílio é positivamente correlacionada aos salários, indicando que jovens de famílias cujo pai tem mais anos de estudo ganham, em média, maiores salários. Os homens recebem mais que as mulheres, em média.

Nas colunas (IV) e (V) são introduzidas ao modelo as variáveis binárias que identificam as Unidades da Federação brasileiras. Essas colunas são desconsideradas na análise já que essas estimações têm poucos graus de liberdade. Por essa razão, foram introduzidas ao modelo as variáveis binárias que identificam as Regiões brasileiras. A partir de então a educação do pai e o sexo dos indivíduos continuam significantes na determinação dos salários, e a cor dos jovens é significativa apenas na ausência da variável de educação medida em anos de estudo. As estimações que não tem a variável de educação indicam que os jovens de cor branca recebem salários maiores que os demais, em média, mas quando a variável quantitativa de educação é introduzida ao modelo, a cor perde significância, e o modelo passa a indicar que os jovens com 12 ou mais anos de estudo ganham, em média, mais que os de nível educacional inferior. Além disso, o fato de se localizarem na região Sudeste tem impacto negativo e significativo na determinação dos salários apenas na presença dessa variável de educação no modelo.

O coeficiente estimado da variável de proficiência indica a elasticidade do salário em relação ao desempenho escolar, já que ambas as variáveis são inseridas ao modelo em logaritmo. Assim, na última coluna da tabela 6, em que a proficiência aparece positivamente relacionada aos salários futuros, o coeficiente indica que a uma variação percentual na nota do teste corresponde 0,57 de variação percentual no salário seis anos depois.

Na tabela 7 são apresentados os resultados da estimação do modelo que verificamos o impacto do desempenho no exame de proficiência em língua portuguesa em 1995 nos salários seis anos depois, em 2001.

Com exceção dos resultados apresentados nas colunas (IV) e (V), que não são levados em conta por motivos já mencionados, e na coluna (VIII), os demais resultados estimados indicam que o desempenho no exame de proficiência em português é positivamente relacionado aos salários futuros. Isso quer dizer que os jovens que tiveram as maiores nota nesse teste recebem os maiores salários no futuro.

TABELA 7 – PROFICIÊNCIA EM PORTUGUÊS (1995) E SALÁRIOS (2001_PNAD)

	Variável Dependente: Logaritmo do Salário Real Horário								
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
PROFICIÊNCIA EM PORTUGUÊS	2,33 0,61	1,47 0,61	1,13 0,48	0,19 0,86	0,45 0,70	0,91 0,54	1,15 0,48	0,83 0,57	1,12 0,51
VIES MIGRAÇÃO		-1,16 0,46	-1,01 0,44	-0,12 0,93	-0,65 0,80	-0,49 0,72	-0,61 0,57	-136,01 190,33	-126,86 160,92
VIES MIGRAÇÃO ²								1,55 2,18	1,42 1,85
VIES MIGRAÇÃO ³								-0,06 0,08	-0,05 0,07
VIES EDUCACIONAL		1,40 0,42	0,17 0,69	0,31 0,77	0,44 0,91	0,62 0,77	0,74 0,78	2,37 4,73	2,39 4,18
VIES EDUCACIONAL ²								-0,20 0,54	-0,19 0,49
VIES EDUCACIONAL ³								0,07 0,18	0,06 0,16
EDUCA_15			0,62 0,22		1,00 0,33		0,70 0,25		0,71 0,26
FREQ. ESCOLA			-0,50 0,24	-0,21 0,31	-0,61 0,25	-0,28 0,23	-0,65 0,22	-0,26 0,25	-0,65 0,24
MEDCH_82 (em log)			0,25 0,07	-0,14 0,28	0,02 0,24	0,25 0,10	0,26 0,09	0,24 0,10	0,25 0,08
HOMEM			0,21 0,04	0,16 0,05	0,23 0,04	0,17 0,04	0,22 0,04	0,17 0,04	0,23 0,04
BRANCA			0,03 0,06	0,26 0,11	0,06 0,13	0,11 0,06	0,00 0,06	0,10 0,07	0,00 0,07
NORDESTE						-0,06 0,09	-0,07 0,07	-0,05 0,10	-0,08 0,08
CENTRO OESTE						0,08 0,07	-0,01 0,07	0,10 0,07	0,00 0,07
SUDESTE						-0,10 0,07	-0,18 0,07	-0,10 0,08	-0,18 0,07
SUL						0,00 0,09	-0,05 0,09	0,01 0,10	-0,05 0,09
UF				SIM	SIM				
CONSTANTE	-12,34 3,45	-6,43 3,56	-4,94 2,89	0,11 5,16	-1,20 4,20	-4,12 3,39	-5,40 2,91	35,48 55,51	31,91 46,55
R ²	0,225	0,344	0,684	0,741	0,807	0,683	0,726	0,687	0,731
No. OBSERVAÇÕES	72	72	70	70	70	70	70	70	70

Analisando o impacto das outras variáveis no salário quando controlamos o modelo pelo resultado da proficiência em língua portuguesa, observamos que na coluna (II), além da proficiência, as variáveis de correção do viés são significantes. A variável de correção do viés de migração tem impacto negativo, enquanto a variável de correção do viés educacional relaciona-se positivamente aos salários futuros. Já na coluna (III) apenas a variável de correção do viés de migração é significativa. Nas demais estimações essas variáveis perdem significância na determinação dos salários.

Nas estimações em que utilizamos a variável de educação, medida em anos de estudo, como variável explicativa do modelo, além dessa variável apresentar impacto positivo e significativo nos salários futuros, a frequência escolar também é significativa, mas com impacto negativo. Assim, os resultados mostram que os salários dos jovens com 12 ou mais anos de estudo são, em média, maiores que daqueles com nível educacional inferior, e que os jovens que em 2001 freqüentavam a escola ganhavam menos em média. Na ausência do controle por anos de estudo a frequência escolar perde significância para explicar os salários.

Nas estimações apresentadas nessa tabela a variável de educação do pai é positiva e significativamente relacionada aos salários, e a variável de sexo indica que os homens ganham mais que as mulheres. Na presença da variável quantitativa de educação, as estimações indicam que os trabalhadores da Região Sudeste ganham menos que os da região Norte. Por fim, a elasticidade do salário futuro à proficiência apresentada na última coluna indica uma variação mais que proporcional dos salários a variação no desempenho escolar.

A tabela 8 apresenta os resultados da estimação do modelo de pseudo-painel da equação (1), em que, assim como na tabela 6, estimamos o impacto do desempenho no exame de proficiência em matemática nos salários, só que dessa vez cinco anos depois, em 2000 (dados do Censo Demográfico).

Assim como ocorreu com os dados de salário de 2001 da PNAD, as colunas (I) a (III) dessa tabela mostram que os salários dos jovens dessa geração aos 22 e 23 anos de idade são determinados pelo desempenho no exame de proficiência em matemática da mesma ao final do ensino médio aos 17 e 18 anos de idade. Ao contrário dos resultados apresentados na tabela 6, ao inserirmos ao modelo da variável de educação, medida em anos de estudo, a proficiência perde significância na determinação dos salários, mas na ausência dessa variável quantitativa, o desempenho escolar e o salário futuro são positivamente relacionados.

Nas estimações apresentadas nas colunas (II), (VI) e (VII) as duas variáveis de correção do viés do modelo, de migração e educacional, são significantes e se relacionam negativamente e positivamente ao salário, respectivamente. Com relação às outras variáveis explicativas, com exceção da variável que controla pelo nível educacional dos jovens, as demais variáveis são significantes para explicar os salários. A educação do chefe do domicílio é positivamente correlacionada aos salários futuros, indicando que jovens provenientes de famílias cujo pai tem, em média, mais anos de estudo tendem a ganhar maiores salários, em média. Os jovens que freqüentam a escola ganham mais, assim como os brancos e os homens quando comparados a seus pares. As regiões onde trabalham também são significantes para determinar os salários.

TABELA 8 – PROFICIÊNCIA EM MATEMÁTICA (1995) E SALÁRIOS (2000_CENSO)

	Variável Dependente: Logaritmo do Salário Real Horário								
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
PROFICIÊNCIA EM MATEMÁTICA	1,18	1,07	0,21	0,14	0,13	0,22	0,19	0,22	0,19
	0,27	0,20	0,11	0,17	0,15	0,12	0,12	0,12	0,12
VIES MIGRAÇÃO		-1,37	-1,07	-0,02	-0,38	-0,90	-1,03	-65,12	-47,81
		0,27	0,21	0,40	0,46	0,28	0,27	78,60	76,61
VIES MIGRAÇÃO ²								0,71	0,50
								0,89	0,87
VIES MIGRAÇÃO ³								-0,03	-0,02
								0,03	0,03
VIES EDUCACIONAL		1,39	0,39	0,60	0,35	0,73	0,54	-1,47	-1,14
		0,31	0,25	0,29	0,25	0,22	0,24	1,26	1,14
VIES EDUCACIONAL ²								0,30	0,23
								0,18	0,15
VIES EDUCACIONAL ³								-0,10	-0,08
								0,06	0,06
EDUCA_15			0,24		0,54		0,42		0,38
			0,25		0,31		0,31		0,30
FREQ. ESCOLA			0,01	0,06	0,02	0,24	0,20	0,25	0,21
			0,11	0,13	0,12	0,09	0,09	0,09	0,09
MEDCH_82 (em log)			0,27	0,09	0,07	0,18	0,14	0,18	0,15
			0,03	0,08	0,08	0,05	0,06	0,05	0,06
HOMEM			0,15	0,15	0,17	0,16	0,18	0,16	0,17
			0,02	0,02	0,02	0,02	0,02	0,02	0,02
BRANCA			0,07	0,14	0,10	0,08	0,06	0,08	0,06
			0,03	0,05	0,06	0,02	0,03	0,02	0,03
NORDESTE						-0,12	-0,16	-0,12	-0,16
						0,04	0,06	0,05	0,06
CENTRO OESTE						-0,07	-0,11	-0,07	-0,10
						0,04	0,05	0,04	0,05
SUDESTE						-0,16	-0,17	-0,17	-0,17
						0,04	0,04	0,04	0,04
SUL						-0,06	-0,08	-0,08	-0,09
						0,04	0,04	0,04	0,05
UF				SIM	SIM				
CONSTANTE	-5,81	-4,02	0,22	-0,05	0,29	0,10	0,45	19,41	14,88
	1,52	1,22	0,70	0,91	0,90	0,80	0,77	22,87	22,25
R ²	0,222	0,511	0,858	0,950	0,955	0,890	0,895	0,896	0,900
No. OBSERVAÇÕES	99	99	95	95	95	95	95	95	95

Na tabela 9 são apresentados os resultados da estimação do modelo em que a variável dependente, o salário em 2000, é explicada pelo desempenho no exame de proficiência em língua portuguesa realizado em 1995.

Os resultados das estimações mostram que o exame de proficiência em língua portuguesa é positivamente relacionado aos salários. Assim, a geração de jovens que tiveram as maiores nota nesse teste recebe os maiores salários no futuro. A elasticidade estimada do salário futuro à proficiência de acordo com o modelo apresentado na última coluna é de 0,42 .

TABELA 9 – PROFICIÊNCIA EM PORTUGUÊS (1995) E SALÁRIOS (2000_CENSO)

Variável Dependente: Logaritmo do Salário Real Horário									
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
PROFICIÊNCIA EM PORTUGUÊS	1,65	1,00	0,52	0,11	0,06	0,47	0,37	0,51	0,42
	0,39	0,31	0,18	0,20	0,21	0,15	0,15	0,15	0,15
VIES MIGRAÇÃO		-1,37	-0,93	0,39	0,05	-0,77	-0,90	-96,97	-68,24
		0,33	0,25	0,34	0,40	0,32	0,30	86,11	86,21
VIES MIGRAÇÃO ²								1,06	0,73
								0,98	0,98
VIES MIGRAÇÃO ³								-0,04	-0,03
								0,04	0,04
VIES EDUCACIONAL		1,40	0,59	0,30	0,12	0,84	0,66	-0,37	0,05
		0,30	0,26	0,21	0,22	0,23	0,25	1,68	1,52
VIES EDUCACIONAL ²								0,17	0,10
								0,21	0,19
VIES EDUCACIONAL ³								-0,06	-0,04
								0,08	0,07
EDUCA_15			0,27		0,47		0,39		0,39
			0,24		0,29		0,28		0,28
FREQ. ESCOLA			-0,02	-0,08	-0,10	0,21	0,17	0,20	0,16
			0,12	0,12	0,11	0,10	0,11	0,10	0,11
MEDCH_82 (em log)			0,27	0,06	0,05	0,20	0,16	0,21	0,17
			0,03	0,07	0,06	0,05	0,06	0,05	0,06
HOMEM			0,17	0,15	0,17	0,18	0,19	0,18	0,19
			0,02	0,01	0,02	0,02	0,02	0,02	0,02
BRANCA			0,06	0,19	0,15	0,07	0,05	0,07	0,05
			0,03	0,04	0,05	0,02	0,03	0,03	0,03
NORDESTE				-0,12	-0,10	-0,10	-0,14	-0,10	-0,14
				0,04	0,04	0,05	0,06	0,05	0,06
CENTRO OESTE						-0,05	-0,09	-0,05	-0,09
						0,04	0,05	0,04	0,05
SUDESTE						-0,15	-0,16	-0,15	-0,16
						0,04	0,04	0,04	0,04
SUL						-0,05	-0,07	-0,06	-0,08
						0,04	0,04	0,04	0,05
UF				SIM	SIM				
CONSTANTE	-8,49	-3,62	-1,66	-0,11	0,42	-1,46	-0,77	27,17	19,69
	2,22	1,84	1,15	1,15	1,27	1,00	1,03	25,05	25,02
R ²	0,177	0,442	0,865	0,957	0,960	0,893	0,898	0,898	0,902
No. OBSERVAÇÕES	97	97	93	93	93	93	93	93	93

Com relação ao impacto nos salários das demais variáveis, nas estimações apresentadas nas colunas (II), (III), (VII) e (VII), as variáveis de correção do viés de migração e educacional são significantes, com sinais negativo e positivo, respectivamente. Assim como na tabela 7, freqüentar a escola em 2000, ser filho de pai mais educado, ser homem e/ou ser de cor branca são características que levam os jovens dessa geração a receberem maiores salários que seus pares. Por outro lado, os trabalhadores das regiões Sudeste e Nordeste ganham menos que os da região Norte, tudo mais constante.

Comparando os resultados das estimações do impacto do desempenho escolar medido pela nota no teste de proficiência em matemática em 1995 nos salários em 2000, cinco anos depois (tabela 8), e em 2001, seis anos depois (tabela 6), observamos que de acordo com os resultados apresentados nas colunas (I) a (III) de ambas as tabelas a proficiência em matemática é significativa e positivamente correlacionada com os salários futuros. Entretanto, o controle do modelo pelas regiões faz com que a proficiência deixe de ser significativa na determinação dos salários na presença da variável de anos de estudo para as estimações com dados do Censo, e na ausência dessa variável nas estimações com dados da PNAD 2001.

Com relação ao impacto das demais variáveis explicativas no salário futuro, os resultados obtidos tanto com os dados do Censo como com os dados da PNAD indicam que nos modelos das colunas (II), (III) e (VII) das tabelas, a variável de controle do viés de migração tem impacto negativo e significativo no salário, e apenas na coluna (II) o impacto positivo da variável de correção do viés educacional é significativo.

O “*background*” familiar, representado pela educação do pai, tem influência significativa no salário, e indica que os trabalhadores cujo pai tem nível educacional mais elevado ganham mais. Além disso, segundo a variável de sexo dos indivíduos, tudo mais constante, os homens são mais bem remunerados, em média, do que as mulheres.

Comparando os resultados que mostram os resultados das regressões que estimam o impacto do desempenho escolar em língua portuguesa ao final do ensino médio nos salários em 2000, cinco anos depois (tabela 9), e em 2001, seis anos depois (tabela 7), observamos que, com exceção de uma das colunas consideradas, a proficiência em português aparece como significativa e positivamente correlacionada com o salário futuro. De acordo com os modelos estimados a geração de jovens com melhor desempenho em português recebe salários futuros maiores. Além disso, as estimações com esses dois bancos de dados mostram a relação positiva entre a educação do chefe do domicílio quando a geração tem 4 e 5 anos de idade e o salário futuro da mesma, e uma discriminação salarial em favor dos homens.

Portanto, concluímos a partir da observação dessas quatro tabelas, que os salários da geração de jovens que nasceram em 1977 e 1978 são determinados pelo desempenho escolar da mesma, medido pela nota obtida em testes de proficiência.

5.1 Primeiro Estágio

As tabelas 10 e 11 reproduzem os resultados das estimações da equação (2) do primeiro estágio, respectivamente para os dados de proficiência em matemática e em português do SAEB de 1995. Os resultados para o desempenho na prova de matemática indicam que todas as características dos alunos consideradas nas estimações foram significantes na determinação da proficiência dessa disciplina: a variável de *'background'* familiar (educação da mãe) se mostrou positivamente relacionada com a proficiência, homens e brancos apresentam melhores desempenhos que seus pares e a idade é negativamente relacionada à nota de matemática. Porém, apenas os salários dos professores são significantes para explicar o desempenho dos alunos nessa disciplina, indicando que quanto maior o salário dos professores, melhor o desempenho dos alunos no teste de proficiência. Com relação aos diretores, além dos salários, o nível educacional deles é significativo para determinar o desempenho escolar dos alunos nessa disciplina. Diretores mais bem pagos e com mais anos de estudo são responsáveis por um desempenho melhor dos alunos no exame do SAEB. Por fim, notamos que as escolas da zona urbana têm um desempenho pior, e as da rede privada melhor, quando comparadas a seus pares. A presença de laboratórios de ciências, computadores ou biblioteca não é estatisticamente significativa.

A tabela 11 apresenta os resultados do impacto das características de alunos, professores, diretores e escolas na proficiência em português. Também com relação ao desempenho nessa disciplina todas as características dos alunos são significantes. Assim como ocorre com o desempenho em matemática, as pessoas de mais idade tem resultado inferior na prova, e as pessoas de cor branca e as com mães mais educadas obtêm notas mais altas. Entretanto, as mulheres têm um melhor desempenho que os homens nessa disciplina. Novamente, os salários dos professores são significantes para explicar o desempenho dos alunos. O nível educacional dos professores também foi significativo e positivamente correlacionado a tais notas. As mesmas características dos professores, que têm impacto significativo nas notas dos alunos, são as características dos diretores relevantes no desempenho escolar em língua portuguesa: salários e nível educacional. Ao contrário do resultado no teste de matemática, o desempenho das escolas da rede urbana é melhor em português que o desempenho das escolas da zona rural. A rede privada também tem desempenho melhor nessa disciplina do que a pública. A presença de computadores e laboratórios de ciências é positivamente relacionada às notas do teste.

TABELA 10 – PRIMEIRO ESTÁGIO: FUNÇÃO DE PRODUÇÃO
EDUCACIONAL – MATEMÁTICA

Variável Dependente: Proficiência em Matemática em logaritmo				
	(I)	(II)	(III)	(IV)
IDADE	-0,07	-0,07	-0,07	-0,06
	0,01	0,01	0,01	0,01
HOMEM	0,09	0,08	0,08	0,08
	0,01	0,01	0,01	0,01
BRANCA	0,04	0,04	0,03	0,03
	0,01	0,01	0,01	0,01
EDUCA MÃE	0,08	0,05	0,04	0,02
Colegial	0,01	0,01	0,01	0,01
SALARIO PROF.1		0,05	0,02	0,03
R\$ 308 a 1023		0,01	0,01	0,01
SALARIO PROF.2		0,13	0,09	0,09
R\$ 1024 a 2046		0,01	0,01	0,01
SALARIO PROF.3		0,15	0,11	0,10
maior que R\$ 2046		0,02	0,02	0,02
PROF. MAGIST.1		0,05	0,04	0,02
de 1 a 10 anos		0,03	0,03	0,03
PROF. MAGIST.2		0,06	0,05	0,03
mais de 10 anos		0,03	0,03	0,03
EDUC.PROF.1		0,01	0,00	0,01
Superior		0,01	0,01	0,02
EDUC.PROF.2		0,02	0,01	0,01
Pós-graduação		0,01	0,02	0,02
SALARIO DIRETOR1			0,02	0,04
R\$ 308 a 1023			0,02	0,02
SALARIO DIRETOR2			0,04	0,05
R\$ 1024 a 2046			0,02	0,02
SALARIO DIRETOR3			0,11	0,08
maior que R\$ 2046			0,02	0,02
DIRETOR MAGIST.1			0,07	0,04
de 1 a 10 anos			0,04	0,04
DIRETOR MAGIST.2			0,08	0,06
mais de 10 anos			0,04	0,04
EDUC.DIRETOR1			0,04	0,04
Superior			0,02	0,02
EDUC.DIRETOR2			0,04	0,05
Pós-graduação			0,02	0,02
ZONA URBANA				-0,07
				0,03
REDE PRIVADA				0,10
				0,01
BIBLIOTECA				0,01
				0,01
LAB. CIÊNCIAS				0,00
				0,01
COMPUTADORES				0,02
				0,01
CONSTANTE	6,83	6,63	6,50	6,42
	0,11	0,11	0,12	0,13
R ²	0,18	0,23	0,27	0,33
No. OBSERVAÇÕES	2909	2824	2614	2397

TABELA 11 – PRIMEIRO ESTÁGIO: FUNÇÃO DE PRODUÇÃO
EDUCACIONAL – PORTUGUÊS

Variável Dependente: Proficiência em Português em logaritmo				
	(I)	(II)	(III)	(IV)
IDADE	-0,06	-0,06	-0,06	-0,05
	0,01	0,01	0,01	0,01
HOMEM	-0,02	-0,02	-0,02	-0,01
	0,01	0,01	0,01	0,01
BRANCA	0,02	0,02	0,01	0,01
	0,01	0,01	0,01	0,01
EDUCA MÃE	0,07	0,05	0,04	0,03
Colegial	0,01	0,01	0,01	0,01
SALARIO PROF.1		0,04	0,02	0,02
R\$ 308 a 1023		0,01	0,01	0,01
SALARIO PROF.2		0,08	0,05	0,04
R\$ 1024 a 2046		0,01	0,01	0,01
SALARIO PROF.3		0,11	0,07	0,06
maior que R\$ 2046		0,02	0,02	0,02
PROF. MAGIST.1		0,02	0,01	-0,02
de 1 a 10 anos		0,03	0,03	0,03
PROF. MAGIST.2		0,03	0,02	-0,01
mais de 10 anos		0,03	0,03	0,03
EDUC.PROF.1		0,07	0,04	0,04
Superior		0,01	0,02	0,02
EDUC.PROF.2		0,06	0,03	0,02
Pós-graduação		0,02	0,02	0,02
SALARIO DIRETOR1			0,05	0,06
R\$ 308 a 1023			0,02	0,02
SALARIO DIRETOR2			0,07	0,07
R\$ 1024 a 2046			0,02	0,02
SALARIO DIRETOR3			0,12	0,10
maior que R\$ 2046			0,02	0,02
DIRETOR MAGIST.1			-0,02	-0,02
de 1 a 10 anos			0,05	0,05
DIRETOR MAGIST.2			-0,01	-0,02
mais de 10 anos			0,05	0,05
EDUC.DIRETOR1			0,07	0,06
Superior			0,02	0,02
EDUC.DIRETOR2			0,07	0,06
Pós-graduação			0,02	0,02
ZONA URBANA				0,08
				0,03
REDE PRIVADA				0,03
				0,01
BIBLIOTECA				0,00
				0,01
LAB. CIÊNCIAS				0,02
				0,01
COMPUTADORES				0,02
				0,01
CONSTANTE	6,70	6,51	6,45	6,30
	0,12	0,12	0,13	0,14
R ²	0,08	0,12	0,16	0,17
No. OBSERVAÇÕES	2987	2907	2686	2450

Concluímos que investimentos em educação, como remuneração e qualificação de docentes e diretores, assim como em infra-estrutura escolar são importantes para proporcionar melhor desempenho dos alunos nos testes de proficiência em ambas as disciplinas. Além disso, observa o impacto positivo do *background* familiar, caracterizado pelo nível educacional da mãe, no desempenho escolar dos alunos.

Os resíduos da estimação do primeiro estágio de cada disciplina (coluna (IV) das tabelas 10 e 11) são agregados pelos vetores características e utilizados no segundo estágio. Esses resíduos representam os fatores que influenciam na nota dos testes de matemática e de português que não são explicados pelas variáveis consideradas.

5.2 Segundo Estágio

Nesse segundo estágio inserimos as equações do modelo de pseudo-painel os resíduos das regressões do primeiro estágio. Esses resíduos representam os determinantes da proficiência que não são considerados nas estimações do primeiro estágio⁶.

A tabela 12 reproduz os resultados da estimação da equação (3) do modelo, em que verificamos o impacto do desempenho no exame de proficiência em matemática realizado pelo SAEB em 1995 nos salários dos jovens dessa geração em 2001. Essa estimação é ponderada pelos resíduos da regressão do primeiro estágio em que foram analisados os determinantes da proficiência em matemática. Esses resíduos representam os determinantes da proficiência nessa disciplina que não são considerados nessas estimações do primeiro estágio.

As estimações apresentadas em todas as colunas dessa tabela (exceção das colunas (IV) e (V) que são desconsideradas da análise por apresentarem poucos graus de liberdade) mostram uma relação positiva e significativa entre o resultado do teste de proficiência em matemática de uma geração e os salários futuros da mesma. Isto quer dizer, que melhores desempenhos no exame de proficiência ao final do ensino médio levam obtenção de maiores salários anos depois. Com relação à nova variável inserida no modelo, nota-se uma relação negativa e significativa entre os resíduos do primeiro estágio e os salários, indicando que quanto maior os determinantes da proficiência em matemática em 1995 que não são considerados na estimação do primeiro estágio menor os salários em 2001. Entretanto, tal relação deixa de ser significativa com a inclusão das variáveis binárias de região.

⁶ Os resultados apresentados em negrito nessas tabelas indicam as variáveis estatisticamente significantes à 10% na determinação da variável dependente.

Com relação ao coeficiente da variável de qualidade da educação, o coeficiente do exame de proficiência, que mede a elasticidade, indica que a uma variação percentual na nota do teste corresponde 0,98 de variação percentual no salário seis anos depois. Assim, de acordo com esse modelo, a variação no salário futuro é quase proporcional a variação no desempenho escolar. O valor dessa elasticidade é 70% maior que a elasticidade estimada no modelo sem a variável que representa os determinantes da proficiência que não são considerados nas estimações do primeiro estágio.

TABELA 12 – SEGUNDO ESTÁGIO: PROFICIÊNCIA EM MATEMÁTICA (1995) E SALÁRIOS (2001_PNAD)

	Variável Dependente: Logaritmo do Salário Real Horário								
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
PROFICIÊNCIA EM MATEMÁTICA	2,16 0,33	1,93 0,27	1,04 0,37	0,75 0,67	0,57 0,55	0,93 0,39	1,05 0,34	0,90 0,41	0,98 0,40
RESÍDUOS A+P+D+E	-1,82 0,53	-1,58 0,44	-0,94 0,45	-0,01 0,68	0,12 0,65	-0,43 0,48	-0,49 0,49	-0,45 0,49	-0,43 0,50
VIES MIGRAÇÃO		-0,97 0,39	-0,88 0,44	-0,48 0,85	-1,18 0,77	-0,88 0,69	-1,18 0,55	-151,92 199,74	-198,14 169,49
VIES MIGRAÇÃO ²								1,76 2,28	2,26 1,96
VIES MIGRAÇÃO ³								-0,07 0,09	-0,09 0,07
VIES EDUCACIONAL		1,50 0,43	-0,43 0,59	-0,94 0,95	-0,79 0,82	-0,15 0,76	-0,01 0,67	-1,08 5,06	0,19 4,32
VIES EDUCACIONAL ²								0,07 0,55	-0,06 0,46
VIES EDUCACIONAL ³								-0,01 0,18	0,03 0,15
EDUCA_15			0,69 0,18		1,05 0,31		0,80 0,20		0,81 0,20
FREQ. ESCOLA			-0,23 0,27	-0,03 0,39	-0,48 0,32	-0,11 0,28	-0,49 0,26	-0,10 0,30	-0,50 0,29
MEDCH_82 (em log)			0,24 0,06	0,02 0,29	0,16 0,25	0,27 0,09	0,27 0,08	0,28 0,09	0,27 0,08
HOMEM			0,07 0,06	0,04 0,10	0,14 0,08	0,04 0,07	0,09 0,06	0,04 0,07	0,11 0,06
BRANCA			0,03 0,06	0,26 0,13	0,09 0,12	0,12 0,06	0,01 0,06	0,11 0,07	0,00 0,07
NORDESTE						-0,01 0,09	-0,05 0,08	0,02 0,09	-0,03 0,08
CENTRO OESTE						0,02 0,08	-0,11 0,08	0,05 0,08	-0,08 0,09
SUDESTE						-0,08 0,09	-0,17 0,08	-0,07 0,10	-0,17 0,09
SUL						0,04 0,09	-0,04 0,09	0,06 0,10	-0,03 0,10
UF				SIM	SIM				
CONSTANTE	-11,29 1,86	-9,21 1,64	-4,50 2,33	-2,79 3,97	-1,46 3,35	-3,80 2,47	-4,21 2,07	39,46 57,90	53,03 48,78
R ²	0,446	0,587	0,732	0,783	0,857	0,702	0,760	0,709	0,768
No. OBSERVAÇÕES	65	65	63	63	63	63	63	63	63

Com relação ao comportamento dos salários em função das demais variáveis do modelo, nas segunda e terceira colunas dessa tabela, ambas as variáveis de correção do viés do modelo são significantes e apresentam os sinais esperados. A variável de correção do viés de migração tem sinal negativo e a variável de correção do viés educacional tem sinal positivo. A variável de educação medida em anos de estudo e a variável de educação do pai têm impactos positivos e significantes, indicando que os jovens com 12 ou mais anos de estudo ganham, em média, mais que os de nível educacional inferior, assim como os filhos de pais com nível educacional mais elevado. O sexo foi significativo apenas nas colunas VII e IX, e a cor do indivíduo na coluna VI. De acordo com essas variáveis, os homens e os brancos ganham salários maiores que seus pares.

Na tabela 13 estão expostos os resultados da estimação do modelo, em que verificamos o impacto do desempenho no exame de proficiência em língua portuguesa realizado pelo SAEB em 1995 nos salários dos jovens da geração que nasceu em 1977 e 1978 em 2001. Essa estimação é ponderada pelos resíduos da regressão do primeiro estágio em que foram analisados os determinantes da proficiência nessa disciplina. Como já mencionado, esses resíduos representam os determinantes dessa proficiência que não são considerados nas estimações do primeiro estágio.

As estimações das três primeiras colunas mostram uma relação significativa e positiva entre o desempenho na prova de português e os salários. Porém, a partir da coluna (IV), a proficiência em português deixa de ser significativa para a determinação dos salários. Na ausência da variável de educação em anos de estudo, os resíduos da estimação do primeiro estágio passam a influenciar positiva e significativamente os salários, o que quer dizer que os salários são influenciados por outras características dos indivíduos que não as consideradas nas estimações de proficiência em português, pelo desempenho escolar nessa disciplina não explicado pelo modelo.

Analisando as demais variáveis, na estimação apresentada na segunda coluna a variável de correção do viés educacional é significativa com sinal positivo e a de correção do viés de migração é significativa com sinal negativo. Já na coluna seguinte, apenas a variável de correção do viés de migração continua significativa na determinação do salário futuro. A educação do pai é significativa com impacto positivo, assim como o nível educacional e o sexo do indivíduo. Os jovens com pais com mais anos de estudos completos, os jovens mais educados e os homens ganham mais que seus pares.

TABELA 13 – SEGUNDO ESTÁGIO: PROFICIÊNCIA EM PORTUGUÊS (1995)
E SALÁRIOS (2001_PNAD)

	Variável Dependente: Logaritmo do Salário Real Horário								
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
PROFICIÊNCIA EM PORTUGUÊS	2,64 0,69	1,51 0,69	0,57 0,53	-1,44 1,02	-0,97 0,71	0,29 0,65	0,62 0,57	0,26 0,70	0,62 0,62
RESÍDUOS A+P+D+E	-0,49 0,71	-0,18 0,70	0,73 0,62	1,77 0,73	1,28 0,50	1,06 0,62	0,70 0,59	1,08 0,66	0,69 0,60
VIES MIGRAÇÃO		-1,04 0,47	-1,03 0,42	-0,61 0,84	-1,16 0,66	-0,92 0,72	-0,98 0,60	-96,23 195,87	-113,08 163,67
VIES MIGRAÇÃO ²								1,09 2,25	1,26 1,89
VIES MIGRAÇÃO ³								-0,04 0,09	-0,05 0,07
VIES EDUCACIONAL		1,45 0,46	0,27 0,76	-0,10 0,89	0,43 0,97	0,49 0,85	0,81 0,83	-1,06 4,96	0,22 4,08
VIES EDUCACIONAL ²								0,12 0,57	0,03 0,50
VIES EDUCACIONAL ³								-0,02 0,19	0,00 0,17
EDUCA_15			0,58 0,23		1,08 0,30		0,70 0,24		0,72 0,26
FREQ. ESCOLA			-0,42 0,28	-0,03 0,40	-0,42 0,30	-0,30 0,30	-0,65 0,28	-0,32 0,32	-0,67 0,30
MEDCH_82 (em log)			0,27 0,07	-0,18 0,28	0,05 0,23	0,28 0,09	0,29 0,09	0,27 0,10	0,28 0,09
HOMEM			0,21 0,04	0,14 0,05	0,22 0,05	0,18 0,05	0,24 0,04	0,18 0,05	0,24 0,04
BRANCA			0,04 0,07	0,36 0,11	0,09 0,13	0,14 0,07	0,01 0,07	0,13 0,08	0,01 0,07
NORDESTE						-0,08 0,10	-0,08 0,08	-0,08 0,11	-0,09 0,09
CENTRO OESTE						-0,01 0,10	-0,08 0,09	0,01 0,11	-0,06 0,10
SUDESTE						-0,14 0,08	-0,20 0,07	-0,14 0,09	-0,20 0,08
SUL						-0,03 0,11	-0,06 0,09	-0,02 0,11	-0,06 0,09
UF				SIM	SIM				
CONSTANTE	-14,08 3,94	-6,82 4,00	-1,80 3,04	9,88 5,99	7,24 4,13	-0,19 3,92	-2,07 3,40	27,60 56,45	31,04 46,86
R ²	0,244	0,347	0,689	0,761	0,832	0,688	0,731	0,691	0,735
No. OBSERVAÇÕES	65	65	63	63	63	63	63	63	63

A tabela 14 reproduz os resultados da estimação da equação (3) do modelo, em que estimamos o impacto do desempenho no exame de proficiência em matemática realizado pelo SAEB em 1995 nos salários dos jovens dessa geração cinco anos depois, em 2000. Essa estimação é ponderada pelos resíduos da regressão do primeiro estágio em que foram analisados os determinantes da proficiência em matemática.

TABELA 14 – SEGUNDO ESTÁGIO: PROFICIÊNCIA EM MATEMÁTICA (1995)
E SALÁRIOS (2000_CENSO)

	Variável Dependente: Logaritmo do Salário Real Horário										
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)
PROFICIÊNCIA EM MATEMÁTICA	2,00 0,33	1,79 0,25	0,63 0,19	0,40 0,17	0,36 0,19	0,56 0,15	0,55 0,16	0,67 0,19	0,66 0,19		
RESÍDUOS A+P+D+E	-1,59 0,48	-1,31 0,39	-0,42 0,23	-0,05 0,21	-0,09 0,23	-0,17 0,21	-0,19 0,21	-0,28 0,23	-0,28 0,23	0,13 0,19	0,10 0,20
VIES MIGRAÇÃO		-1,01 0,21	-0,92 0,22	0,00 0,36	-0,57 0,47	-0,92 0,29	-0,98 0,27	-91,04 88,97	-87,21 89,07	-1,12 0,28	-24,19 89,08
VIES MIGRAÇÃO ²								1,04 1,02	0,99 1,02		0,24 1,02
VIES MIGRAÇÃO ³								-0,04 0,04	-0,04 0,04		-0,01 0,04
VIES EDUCACIONAL		1,52 0,27	0,46 0,26	0,67 0,30	0,30 0,31	0,64 0,23	0,56 0,26	-1,17 1,56	-1,11 1,53	0,51 0,27	-0,22 1,65
VIES EDUCACIONAL ²								0,27 0,20	0,26 0,19		0,13 0,20
VIES EDUCACIONAL ³								-0,10 0,07	-0,10 0,07		-0,05 0,07
EDUCA_15			0,13 0,21		0,60 0,37		0,17 0,23		0,07 0,22	0,22 0,24	0,20 0,24
FREQ. ESCOLA			0,00 0,10	-0,06 0,14	-0,05 0,12	0,21 0,10	0,19 0,10	0,23 0,11	0,22 0,11	0,26 0,11	0,27 0,12
MEDCH_82 (em log)			0,27 0,04	0,14 0,09	0,15 0,08	0,21 0,05	0,20 0,05	0,21 0,05	0,20 0,05	0,19 0,05	0,20 0,05
HOMEM			0,11 0,03	0,12 0,02	0,15 0,03	0,13 0,02	0,14 0,03	0,12 0,03	0,13 0,03	0,20 0,02	0,19 0,02
BRANCA			0,06 0,03	0,12 0,05	0,07 0,06	0,06 0,02	0,05 0,03	0,04 0,03	0,04 0,03	0,06 0,03	0,06 0,03
NORDESTE						-0,09 0,04	-0,10 0,05	-0,07 0,04	-0,08 0,05	-0,12 0,05	-0,12 0,05
CENTRO OESTE						-0,08 0,04	-0,10 0,05	-0,09 0,05	-0,10 0,05	-0,10 0,05	-0,10 0,06
SUDESTE						-0,15 0,04	-0,15 0,04	-0,17 0,04	-0,17 0,05	-0,17 0,05	-0,18 0,05
SUL						-0,04 0,04	-0,05 0,04	-0,05 0,04	-0,05 0,05	-0,07 0,05	-0,08 0,05
UF				SIM	SIM						
CONSTANTE	-10,45 1,89	-8,46 1,50	-2,25 1,15	-1,52 0,95	-0,88 1,03	-1,78 0,95	-1,65 0,95	23,63 25,47	22,64 25,51	1,52 0,28	8,75 25,93
R ²	0,470	0,722	0,889	0,961	0,965	0,918	0,918	0,923	0,923	0,909	0,911
No. OBSERVAÇÕES	82	82	78	78	78	78	78	78	78	78	78

Assim como aconteceu com os dados de salário de 2001 (PNAD), todas as colunas da tabela 14 (inclusive as colunas do modelo ponderado pelas dummies de Unidade da Federação), com dados de salários do Censo Demográfico de 2000, mostram uma relação positiva e significativa entre o resultado do teste de proficiência em matemática e os salários futuros. O que significa que há uma relação positiva entre a qualidade da educação e os salários futuros no Brasil. Com relação a variável que representa os resíduos do primeiro

estágio do modelo, novamente observamos uma relação negativa e significativa entre os resíduos do primeiro estágio e os salários em 2000 nas três primeiras colunas. Tal relação deixa de ser significativa a partir de então.

Com relação a elasticidade do salário à variação no desempenho escolar, o coeficiente estimado da proficiência na estimação apresentada na coluna (IX) indica que a cada variação percentual na nota de matemática, os salários futuros sofrem variação percentual de 0,66 na mesma direção.

Os resultados apresentados nas colunas (X) e (XI) reproduzem as estimações dos modelos apresentados nas colunas (VII) e (IX), respectivamente, na ausência da variável de qualidade da educação. Comparando as estimações, ao inserirmos a variável de qualidade da educação ao modelo, a variável de educação medida em anos de estudo tem seu impacto reduzido no salário, assim como seu nível de significância, apesar dos anos de estudo não serem estatisticamente diferente de zero em nenhum dos casos comparados.

Com relação ao comportamento das demais variáveis na determinação dos salários, as variáveis de correção do viés de migração e educacional são significantes, com sinais negativo e positivo respectivamente para explicar os salários nas colunas (II), (III), (VI) e (VII). As variáveis de frequência à escola, educação do pai e sexo têm impactos estatisticamente significantes nos salários. Os jovens que em 2000 freqüentam à escola ganham mais que os que não freqüentavam, os jovens com pais mais educados recebem maiores salários, e os homens têm salários mais altos que as mulheres. O fato de ser branco não é significativo para explicar o diferencial de salários nos modelos apresentados nas colunas (VIII) e (IX).

Na tabela 15 estão os resultados da estimação do modelo em que verificamos o impacto do desempenho no exame de proficiência em língua portuguesa realizado pelo SAEB em 1995 nos salários dos jovens da geração analisada em 2000. O modelo tem os resíduos da regressão do primeiro estágio em que foram analisados os determinantes da proficiência nessa disciplina.

Ao contrário do que apresentou as estimações com os dados de salário de 2001 (PNAD), os resultados das estimações com os dados de salários do Censo Demográfico de 2000 mostram uma relação positiva e significativa entre o resultado do teste de proficiência em língua portuguesa e os salários futuros. Com relação a variável que representa os resíduos do primeiro estágio do modelo, observa-se uma relação negativa e significativa entre os resíduos do primeiro estágio e os salários em 2000 apenas nas duas primeiras colunas dessa

tabela, sendo que nas demais essa variável não é estatisticamente significativa. A elasticidade estimada dos salários futuros é de 0,49 com relação ao desempenho em português.

TABELA 15 – SEGUNDO ESTÁGIO: PROFICIÊNCIA EM PORTUGUÊS (1995)
E SALÁRIOS (2000_CENSO)

	Variável Dependente: Logaritmo do Salário Real Horário										
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)
PROFICIÊNCIA EM PORTUGUÊS	2,62 0,48	1,60 0,39	0,63 0,24	0,15 0,24	-0,02 0,26	0,60 0,20	0,49 0,21	0,64 0,22	0,49 0,25		
RESÍDUOS A+P+D+E	-1,59 0,51	-1,02 0,43	-0,30 0,24	-0,07 0,22	-0,03 0,20	-0,13 0,19	-0,15 0,20	-0,15 0,20	-0,13 0,22	0,12 0,18	0,15 0,18
VIES MIGRAÇÃO		-1,22 0,33	-0,95 0,22	0,31 0,35	-0,27 0,43	-0,96 0,25	-1,09 0,26	-44,90 80,13	-12,97 88,04	-1,24 0,25	48,92 76,69
VIES MIGRAÇÃO ²								0,51 0,92	0,13 1,01		-0,60 0,88
VIES MIGRAÇÃO ³								-0,02 0,03	0,00 0,04		0,02 0,03
VIES EDUCACIONAL		1,43 0,31	0,78 0,25	0,19 0,28	0,00 0,27	0,91 0,23	0,76 0,25	0,40 2,08	0,97 1,98	0,66 0,25	1,32 1,84
VIES EDUCACIONAL ²								0,04 0,24	-0,04 0,23		-0,09 0,21
VIES EDUCACIONAL ³								-0,01 0,08	0,01 0,08		0,03 0,07
EDUCA_15			0,31 0,21		0,70 0,30		0,38 0,26		0,39 0,30	0,49 0,27	0,54 0,28
FREQ. ESCOLA			-0,10 0,11	-0,08 0,13	-0,10 0,11	0,14 0,11	0,11 0,11	0,14 0,11	0,10 0,12	0,15 0,11	0,11 0,12
MEDCH_82 (em log)			0,27 0,04	0,04 0,07	0,05 0,07	0,21 0,05	0,17 0,06	0,21 0,05	0,18 0,06	0,16 0,06	0,16 0,06
HOMEM			0,18 0,02	0,15 0,01	0,17 0,02	0,19 0,02	0,20 0,02	0,19 0,02	0,20 0,02	0,20 0,02	0,20 0,02
BRANCA			0,06 0,03	0,22 0,04	0,14 0,05	0,08 0,03	0,06 0,03	0,08 0,03	0,06 0,03	0,07 0,03	0,08 0,03
NORDESTE						-0,10 0,05	-0,12 0,06	-0,09 0,06	-0,13 0,07	-0,15 0,06	-0,17 0,06
CENTRO OESTE						-0,08 0,05	-0,12 0,06	-0,08 0,05	-0,11 0,06	-0,14 0,06	-0,14 0,06
SUDESTE						-0,15 0,04	-0,16 0,04	-0,15 0,05	-0,15 0,04	-0,17 0,04	-0,16 0,04
SUL						-0,05 0,04	-0,06 0,05	-0,05 0,05	-0,06 0,05	-0,08 0,05	-0,08 0,05
UF				SIM	SIM						
CONSTANTE	-14,01 2,75	-7,18 2,24	-2,24 1,39	-0,24 1,44	1,14 1,60	-2,00 1,22	-1,27 1,27	10,41 22,93	2,45 24,82	1,66 0,26	-12,24 22,05
R ²	0,307	0,515	0,882	0,967	0,971	0,911	0,915	0,912	0,915	0,909	0,911
No. OBSERVAÇÕES	82	82	78	78	78	78	78	78	78	78	78

Assim como na tabela 14, os resultados apresentados nas colunas (X) e (XI) reproduzem as estimações dos modelos apresentados nas colunas (VII) e (IX), respectivamente, na ausência da variável de qualidade da educação. Comparando as

estimações, ao inserirmos a variável de qualidade da educação ao modelo, a variável de educação medida em anos de estudo deixa de ser significativa na determinação dos salários futuros. Esse resultado foi obtido por muitos trabalhos internacionais sobre esse tema.

Observando a relação das outras variáveis e os salários, as variáveis de correção do viés de migração e educacional são significantes na determinação dos salários nas colunas (II), (III), (VI) e (VII). Esses resultados são idênticos aos apresentados na tabela anterior em que se regressou os salários contra o desempenho em matemática.

A variável de nível educacional aparece como sinal positivo e é significativa apenas na coluna (V), mas os sinais apresentados são opostos aos esperados. A variável de frequência à escola passa a ser estatisticamente significativa a partir da inclusão das dummies de região ao modelo. Já as variáveis de educação do pai, sexo e cor têm impactos estatisticamente significantes nos salários sempre. Pessoas com pais mais educados têm maiores os salários, os homens têm salários mais altos que as mulheres, e os brancos recebem mais que os demais.

Comparando as tabelas 12 e 14, que utilizam os dados de proficiência em matemática e de salários em 2001 e 2000, respectivamente, observamos em ambas que o desempenho dos jovens na prova de matemática ao final do ensino médio em 1995, quando eles têm 17 e 18 anos de idade, é estatisticamente significativo para explicar os salários dessa geração seis e cinco anos depois. A relação entre a qualidade do ensino, medida pela nota no teste, e os salários futuros é positiva, indicando que aqueles que têm notas mais altas recebem salários maiores no futuro. As elasticidades estimadas, entretanto, são diferentes. Com dados da PNAD a elasticidade do salário com relação ao desempenho escolar é quase 50% maior que no modelo que usa os dados do Censo. A relação negativa e significativa entre os resíduos do primeiro estágio e os salários aparece em ambas as tabelas nas colunas (I) a (III). A variável de correção do viés de migração é significativa com sinal negativo nas colunas (II) a (III) e (VII), e a de correção do viés educacional tem influencia significativa de acordo com ambas as tabelas apenas na coluna (II). A educação do chefe do domicílio aparece com impacto significativo e positivo nos salários em todas as colunas das suas tabelas analisadas.

Entretanto, comparando as estimações das tabelas 13 e 15 em que analisamos a relação entre o desempenho escolar em língua portuguesa e os salários dos jovens brasileiros, observamos diferenças grandes nos resultados. Ambas as tabelas, mostram que o impacto do desempenho em português é significativo e positivo apenas nas colunas (I) a (III). A educação do chefe do domicílio e o sexo do indivíduo aparecem com impactos significantes nos salários, indicando que as pessoas com pais mais educados e os homens ganham mais.

Em resumo, essas quatro tabelas, que reproduzem os resultados das estimações do segundo estágio, indicam que quando ponderadas pelos resíduos do primeiro estágio, a relação entre o desempenho na prova de matemática do SAEB ao final do ensino médio e os salários futuros é positiva e significativa. Já com relação ao desempenho em língua portuguesa os resultados são contraditórios. Com os dados de salários da PNAD de 2001, a relação entre o desempenho escolar em língua portuguesa e os salários dos jovens brasileiros não é estatisticamente significativa, mas é positiva e significativa a relação entre os resíduos da estimação dos determinantes do desempenho escolar em português e os salários desses jovens. Entretanto, com os dados de salários do Censo Demográfico de 2000, a relação entre o desempenho escolar em língua portuguesa e os salários dos jovens é positiva e significativa.

5.3 Interações

Nesta parte do trabalho o objetivo é verificar se há variação suficiente entre as células de sexo, cor e Unidade da Federação para estimar a proficiência e os salários depois de controlada essas variáveis. Essa verificação é feita de duas formas: pela análise do coeficiente de determinação, R^2 , e pelo teste F de significância conjunta dos parâmetros.

Numa regressão múltipla o coeficiente de determinação vai aumentar (diminuir) quando variáveis explicativas são inseridas (deletadas) do modelo. Isso ocorre, pois acrescentando uma variável ao modelo a Soma dos Quadrados dos Erros diminui.

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT}$$
, em que $SQT = \sum (y_t - \bar{y})^2$ é a soma de quadrados total, $SQR = \sum (\hat{y}_t - \bar{y})^2$ é a soma de quadrados da regressão, e $SQE = \sum \hat{e}_t^2$ é a soma de quadrados dos erros. $SQT = SQR + SQE \Rightarrow \sum (y_t - \bar{y})^2 = \sum (\hat{y}_t - \bar{y})^2 + \sum \hat{e}_t^2$

O Teste F da significância conjunta dos parâmetros do modelo é utilizado para verificar se, em um modelo de regressão múltipla, os parâmetros conjuntamente afetam a variável dependente. Para um modelo com k variáveis exógenas, o modelo não restrito é:

$$y_t = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + e_t, \text{ cuja hipótese nula é } b_1 = b_2 = \dots = b_k = 0$$

A hipótese nula afirma que todos os coeficientes são iguais a zero, e que o modelo que deve ser usado é o modelo restrito. A hipótese alternativa desse teste é de que, pelo menos, um dos coeficientes é diferente de zero. Sob H_0 o modelo restrito é: $y_t = b_0 + e_t$

A estatística F é: $F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)} \approx F(k, n - k - 1)$, em que R^2 é o coeficiente de determinação da regressão de y_i em x_1, x_2, \dots, x_k .

Se o p-valor é menor que o nível de significância que se está testando, 0,05, rejeita-se H_0 de que os coeficientes são iguais a zero.

Nesse trabalho o Teste F é usado para testar a significância conjunta das interações entre as variáveis de sexo, cor e Unidade da Federação, tendo como variável dependente o logaritmo do salário ou o logaritmo da proficiência.

$$\text{Modelo não restrito: } y_i = a + q_1SEXO + q_2COR + q_3UF_i + b_1x_1 + b_2x_2 + \dots + b_kx_k + e_i$$

$$\text{Em que sexo e cor são variáveis binárias, } SEXO \begin{cases} 1 - Homem \\ 0 - Mulher \end{cases} \text{ e } COR \begin{cases} 1 - Branco \\ 0 - Outros \end{cases}$$

UF_i representa as dummies de Unidade da Federação, $i=1, 2, \dots, 27$ e x'_s são as interações.

$$\text{Hipótese Nula: } b_1 = b_2 = \dots = b_k = 0.$$

$$\text{Sob } H_0 \text{ o modelo restrito é: } y_i = a + q_1SEXO + q_2COR + q_3UF_i + e_i$$

As tabelas abaixo mostram os resultados dos coeficientes de determinação para os modelos restrito e não restrito e do teste F de Wald que compara esses modelos.

TABELA 16 – COEFICIENTE DE DETERMINAÇÃO E TESTE F DE SIGNIFICÂNCIA CONJUNTA:
LOGARITMO DA PROFICIÊNCIA EM MATEMÁTICA (1995)

Variável Dependente: Logaritmo da Proficiência em Matemática - 1995			Teste F de significância conjunta dos parâmetros (Interações)	
DUMMY DE SEXO	SIM	SIM	Hipótese Nula: $F(79, 3672) = 1,83$	
DUMMY DE COR	SIM	SIM	INTERAÇÕES = 0 Prob > F = 0,0000	
DUMMIES DE UF	SIM	SIM		
INTERAÇÕES	NÃO	SIM		
R^2	0,151	0,183		
No. OBSERVAÇÕES	3780	3780		
VARIAÇÃO % do R^2	21,26%			

Os resultados da tabela 16 têm como variável dependente o logaritmo da nota da proficiência em matemática obtida em 1995 no exame realizado pelo INEP. De acordo com os

resultados o coeficiente de determinação do modelo não restrito, com as interações entre os vetores características, *sexo*, *cor* e *UF*, é 21% maior do que o do modelo restrito. Além disso, o teste de significância conjunta dos coeficientes apresenta um p-valor igual à zero, o que leva a rejeição da hipótese nula. Isto é, as interações têm importância na determinação da proficiência em matemática em 1995.

A tabela 17 apresenta os resultados das estimações que têm o logaritmo da nota da proficiência em língua portuguesa obtida em 1995. O coeficiente de determinação do modelo não restrito é mais de 41% maior que o do modelo restrito. O teste de significância conjunta dos coeficientes apresenta um p-valor igual a 0.0576, o que leva a rejeição da hipótese nula à 10% de significância, mas a não rejeição de H_0 à 5%.

TABELA 17 – COEFICIENTE DE DETERMINAÇÃO E TESTE F DE SIGNIFICÂNCIA CONJUNTA:
LOGARITMO DA PROFICIÊNCIA EM PORTUGUÊS (1995)

Variável Dependente: Logaritmo da Proficiência em Português - 1995			Teste F de significância conjunta dos parâmetros (Interações)	
DUMMY DE SEXO	SIM	SIM	Hipótese Nula: $F(79, 3902) = 1,27$	
DUMMY DE COR	SIM	SIM	INTERAÇÕES = 0 Prob > F = 0,0576	
DUMMIES DE UF	SIM	SIM		
INTERAÇÕES	NÃO	SIM		
R^2	0,059	0,084		
No. OBSERVAÇÕES	3860	3860		
VARIAÇÃO % do R^2	41,15%			

As tabelas 18 e 19 apresentam a mesma análise para a determinação dos salários em 2001, com dados da PNAD, e em 2000, com dados do Censo Demográfico, respectivamente. A tabela 18 apresenta uma diferença de quase 27,3% entre o coeficiente de determinação dos modelos restrito e não restrito, e um p-valor para o teste de significância conjunta dos coeficientes é de 0.0082, o que leva a rejeição da hipótese nula à 1%. Isso indica que as interações são significantes na determinação dos salários em 2001. Já nas regressões com os salários de 2000 (tabela 19), notamos que o R^2 é menos de 2% maior no modelo com interações do que no modelo sem interações, mas o p-valor igual a zero leva a rejeição de H_0 à 1%, indicando que as interações têm importância na determinação dos salários em 2000.

TABELA 18 – COEFICIENTE DE DETERMINAÇÃO E TESTE F DE SIGNIFICÂNCIA CONJUNTA:
LOGARITMO DO SALÁRIO REAL HORÁRIO (PNAD 2001)

Variável Dependente: Logaritmo do Salário Real Horário - PNAD 2001			Teste F de significância conjunta dos parâmetros (Interações)	
DUMMY DE SEXO	SIM	SIM	Hipótese Nula: $F(64, 895) = 1,44$	INTERAÇÕES = 0 Prob > F = 0,0082
DUMMY DE COR	SIM	SIM		
DUMMIES DE UF	SIM	SIM		
INTERAÇÕES	NÃO	SIM		
R ²	0,102	0,130		
No. OBSERVAÇÕES	3604	3604		
VARIAÇÃO % do R ²	27,25%			

TABELA 19 – COEFICIENTE DE DETERMINAÇÃO E TESTE F DE SIGNIFICÂNCIA CONJUNTA:
LOGARITMO DO SALÁRIO REAL HORÁRIO (CENSO 2000)

Variável Dependente: Logaritmo do Salário Real Horário - Censo 2000			Teste F de significância conjunta dos parâmetros (Interações)	
DUMMY DE SEXO	SIM	SIM	Hipótese Nula: $F(70, 1113) = 3,89$	INTERAÇÕES = 0 Prob > F = 0,0000
DUMMY DE COR	SIM	SIM		
DUMMIES DE UF	SIM	SIM		
INTERAÇÕES	NÃO	SIM		
R ²	0,092	0,094		
No. OBSERVAÇÕES	165550	165550		
VARIAÇÃO % do R ²	1,85%			

Concluí-se que as interações entre as variáveis de *sexo*, *cor* e *Unidade da Federação* são significantes à 1% na determinação tanto do desempenho no teste de proficiência em matemática realizado em 1995, quanto dos salários em 2001, com dados da PNAD, e em 2000 com dados do Censo Demográfico. Com relação ao desempenho em língua portuguesa as interações entre essas três variáveis são significantes a 10%.

6 CONCLUSÃO

Os resultados desse trabalho mostram que investimentos em educação, como melhores remunerações a professores e diretores, critérios mais rigorosos de seleção desses profissionais, assim como investimentos em infra-estrutura, a fim de melhorar a qualidade da escola, têm como consequência um melhor desempenho dos alunos em exames de proficiência.

Ao analisarmos os determinantes da proficiência escolar em matemática e em língua portuguesa observamos que o desempenho dos alunos em testes padronizados é influenciado pelas características próprias como idade, sexo e cor. Os alunos mais jovens e de cor branca obtêm notas maiores que seus pares em ambas as disciplinas. A diferença fica por conta do sexo desses jovens, os homens apresentam desempenho melhor do que as mulheres em matemática, enquanto as mulheres são mais bem avaliadas do que os homens no exame de língua portuguesa.

O impacto do “*background*” familiar representado pela educação da mãe indicou que os filhos de mães com nível educacional mais elevado apresentam desempenho, em média, melhor nos testes de proficiência do que os filhos de mãe com nível educacional inferior. A influência dos professores e diretores no desempenho escolar dos alunos também é significativa. A remuneração e o nível educacional desses profissionais são positivamente relacionados às notas dos alunos. Por fim, a presença de computadores nas escolas influenciou positivamente o resultado dos testes de proficiência dos alunos.

Esses resultados apontam para a existência de uma relação entre os recursos destinados à educação e a qualidade da mesma no Brasil. Isso corrobora os argumentos internacionais sobre a importância de se destinar mais recursos às escolas, não apenas para aumentar os anos de estudos dos indivíduos, mas para melhorar a qualidade desses anos de estudos.

Além disso, os resultados do segundo estágio do modelo mostram que as notas obtidas por uma geração em testes de proficiência são significantes para explicar os salários futuros da mesma. Assim, confirmamos a relação positiva entre desempenho escolar e salários futuros destacada nos estudos internacionais de Bishop (1989), O'Neill (1990), Murnane et al. (1995), Boissiere et al. (1985) feitos para os Estados Unidos e outros países.

Ao analisarmos as estimações do segundo estágio, tanto com dados da PNAD de 2001 quanto do Censo de 2000, os resultados indicam que a nota obtida no exame de proficiência

em matemática pelos jovens em 1995 das coortes de 1977 e 1978 ao final do ensino médio tem impacto positivo nos salários dessa geração aos 22, 23 e 24 anos de idade.

Com relação ao impacto da proficiência em língua portuguesa, os resultados das regressões com dados do Censo de 2000 confirmam a relação positiva e significativa entre a qualidade da escola medida pela nota no exame de proficiência e os salários futuros. Entretanto, os resultados das estimações utilizando os dados da PNAD de 2001 mostram que as notas obtidas por uma geração nos testes realizados pelo SAEB não são significantes para explicar os salários futuros da mesma. De acordo essa análise os salários são influenciados por outras características da geração que não as consideradas nas estimações dos determinantes da proficiência em língua portuguesa, pelo desempenho escolar não explicado pelo modelo.

A maior representatividade das células do Censo, todas com mais de 30 indivíduos cada, dão maior credibilidade aos resultados que usam esse banco de dados em detrimento dos que usam a PNAD, cujas células usadas tinham pelo menos 10 indivíduos. Portanto, também com relação a nota no teste de proficiência em língua portuguesa, podemos considerar que a nota e o salário futuro se relacionam positivamente.

Por fim, destacamos o fato de que ao incluirmos a variável de qualidade da educação ao modelo controlado apenas pela variável de educação medida em anos de estudo, essa variável quantitativa perde significância para explica os salários.

Concluimos então, que aumentar os recursos destinados à educação no Brasil, a fim de melhorar a qualidade dos anos de estudo dos alunos, é importante e necessário, pois a qualidade da educação tem impactos sócio-econômicos positivos. O investimento em educação está fortemente relacionado ao desempenho escolar dos alunos e, esse desempenho, medido pelas notas obtidas por uma geração em exames de proficiência ao término do ensino médio, afeta os salários a serem recebidos por esta geração quando ela estiver no mercado de trabalho, cinco ou seis anos depois. Assim, as iniciativas de políticas não deveriam focar apenas em quantidade de escolaridade, mas também na sua qualidade como defendido por muitos autores internacionais que analisaram o tema em questão.

7 BIBLIOGRAFIA

BASTOS, V. M. E RIBEIRO, E. P. *Viés de Seleção, retornos à educação e migração no Brasil*. **Anais da Sociedade Brasileira de Econometria**. 2004.

BISHOP, J. *Is the test score decline responsible for the productivity growth decline?*. **American Economics Review**. Vol. 79 (1), p. 178-97, 1989.

BISHOP, J. *Achievement, test score, and relative wage*, in (M.H. Kosters, ed.). **Workers and Their wages**. p. 146-86. Washington, DC: The AEI Press. 1991 .

BOISSIERE, M. X., KNIGHT, J. B., AND SABOT, R. H. *Earning, schooling, ability, and cognitive skill*. **American Economics Review**. Vol. 75 (5), p. 1016-30, 1985.

BROWNING, M., DEATON, A, AND IRISH, M. *A Profitable Approach to Labor Supply and Commodity Demands over the Life Cycle*. **Econometrica**. Vol. 59, p.503-44, 1985.

DAHL, G.B. *Mobility and the Return to Education: Testing a Roy Model with Multiple Markets*. **Econometrica**. Vol 70, no. 6, p. 2367-2420. November, 2002.

HANUSHEK, E. A. *A more complete picture of school resource policies*. **Review of Educational Research**. Vol 66 (2) (Fall), p. 397-409, 1996.

HANUSHEK, E. A. (2003). *The Failure of Inputs-Based Schooling Policies*. **Economic Journal**, 113, p. F64-F98, February, 2003.

HANUSHEK, E. A. AND KIMKO, D. D. *Schooling, labor force quality, and the growth of nations*. **American Economic Review**. vol. 90 (5), p.11894-208, December, 2000.

LISBOA, M. E MENEZES-FILHO, N. *Microeconomia e Sociedade no Brasil*. Rio de Janeiro: **EPGE**, 2001.

MENEZES-FILHO, N. *A Evolução Recente da Educação no Brasil. Tese (Livre Docência)* – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, 113p, 2003.

MENEZES-FILHO, N. A. *Qualidade das Escolas, dos Alunos ou dos Professores? Diferenças nos Resultados do SAEB entre 1995 e 2001, entre Escolas Públicas e Privadas e entre Regiões do Brasil*, Universidade de São Paulo, mimeo, 2004.

MENEZES-FILHO, N. A. *Educação e desigualdade*. In Lisboa, M. B. e Menezes-Filho, N., org. **Microeconomia e sociedade no Brasil**. Rio de Janeiro: Contra Capa Livraria, 2001.

MENEZES-FILHO, N. A. *Microeconometria*. In Lisboa, M. B. e Menezes-Filho, N., org. **Microeconomia e sociedade no Brasil**. Rio de Janeiro: Contra Capa Livraria, 2001b.

MURNANE, R. J., WILLETT, J. B. AND LEVY, F. *The growing importance of cognitive skill in wage determination*. **Review of Economics and Statistics**, vol. 77 (2) p. 251-66, May 1995.

MURPHY, K. M. AND PELTZMAN, S. *School Performance and the Youth Labor Market*. **Journal of Labor Economics**, vol.22 (2), pp.299-325, The University of Chicago. 2004.

O'NEILL, J. *The role of human capital in earnings differences between black and white men*. **Journal of Economic Perspective**, vol.4 (5), p. 869-95, October, 1990.

RIVKIN, S. G. *Black/white differences in schooling and employment*. **Journal of Human Resources**, vol.30 (4) (Fall), pp. 826-52, 1995.

ROY, A. D. *Some Thought on the Distribution of Earnings*. **Oxford Economic Paper**, 3, 135-146. 1951.

Sistema Nacional de Avaliação da Educação Básica (2004) – *Resultados do SAEB 2003 Brasil e Rio de Janeiro* – Brasília, DF. Junho, 2004.