

# Disparidades Regionais ou Educacionais?

## Um Exercício Contrafactual

Angelo José Mont'Alverne Duarte \*      Pedro Cavalcanti Ferreira †

Márcio Antônio Salvato‡

Julho/2003

### Abstract

Este trabalho investiga o impacto da escolaridade sobre a distribuição de renda do trabalho de estados/regiões do Brasil. Usando um método semi-paramétrico, discutido em DiNardo, Fortin & Lemieux (1996), mensuramos o quanto dos diferenciais de renda entre as Regiões Nordeste e Sudeste do Brasil - a mais pobre e a mais rica do país - e entre os Estados do Ceará e São Paulo, podem ser explicados pelas diferenças de escolaridade da população residente. Usando dados da PNAD construímos densidades contrafactuais reponderando a distribuição da região/estado mais pobre pelo perfil de escolaridade da mais rica. Concluimos que: (i) mais de 50% do diferencial de renda é explicado pelo diferencial de escolaridade; (ii) a renda do Nordeste reponderada pela escolaridade do Sudeste equivale a 93% da renda média do Brasil e; (iii) o aumento da escolaridade, mantendo-se a estrutura de salários, agrava a desigualdade de renda nas regiões/estados mais pobres.

*JEL Classification Codes:* C14, I20 e J31

---

\*EPGE/FGV e Banco Central do Brasil, e-mail: aduarte@fgvmail.br

†EPGE/FGV, e-mail: ferreira@fgv.br

‡EPGE/FGV e PUC.Minas, e-mail: salvato@fgvmail.br

# 1 Introdução

A desigualdade de renda existente entre as regiões brasileiras, notadamente entre a Região Sudeste e as Regiões Norte e Nordeste, é objeto permanente de discussões entre acadêmicos, elaboradores de políticas governamentais e membros da classe política. Ainda no século XIX o Imperador D. Pedro II já demonstrava sua preocupação ao afirmar que venderia sua coroa para matar a fome dos nordestinos. Ao longo desses quase 150 anos, vários foram os diagnósticos apresentados e as soluções e programas implementados com objetivo de diminuir essa desigualdade. As políticas governamentais se intensificaram e se tornaram sistemáticas a partir da década de 50 quando, sob a liderança do economista Celso Furtado, foi publicado o documento “Uma Política para o Desenvolvimento do Nordeste” que sugeria a industrialização como forma de diminuir a disparidade entre as rendas das regiões nordeste e sudeste. Esse documento inspirou a criação da SUDENE e da SUDAM na qualidade de órgãos planejadores e executores das políticas de combate às desigualdades regionais, e do BNB e BASA na qualidade de agentes financeiros de desenvolvimento regional. A Constituição Federal de 1988 não foi silente em relação à questão e determinou a alocação de 3% das receitas de impostos sobre a renda e produtos industrializados em programas de financiamento ao setor produtivo nas Regiões Norte, Nordeste e Centro-Oeste.

Apesar da elevada quantidade de órgãos e programas governamentais a nível federal e estadual que se prestam ao desenvolvimento econômico regional, as políticas governamentais implementadas nesses últimos 50 anos se utilizaram fortemente dos mesmos instrumentos: concessão de incentivos fiscais e crédito público subsidiado à iniciativa privada e investimento estatal em infra-estrutura. Estima-se que entre 1989 e 2002 os Fundos Constitucionais de desenvolvimento regional tenham aplicado recursos da ordem de US\$ 10 bilhões. Embora bem sucedidos em algumas dimensões, como a de acelerar o processo de industrialização nas Regiões Nordeste e Norte, essas políticas não foram competentes em transformar os indicadores sociais, reduzir o nível de pobreza e modificar a distribuição de renda da região.

Na última década surgiu uma nova abordagem do problema de desigualdade regional no Brasil. Ela baseia-se na tese de que a baixa renda *per capita* nas Regiões Norte e Nordeste está relacionada com a concentração de indivíduos com baixa escolaridade (capital humano) e baixo capital físico, o que fazem com que suas rendas sejam pequenas. Dessa forma a redução da desigualdade regional confunde-se com o combate à pobreza, e para tal necessita-se de uma política de educação e qualificação profissional e programas de acesso ao crédito.

Pessoa (2000) sugere que há dois enfoques ao problema de desigualdade de renda: o primeiro, e mais importante, está relacionado à diferença de renda per capita entre as regiões; e o segundo, bem menos significativo, se refere à distribuição espacial da produção. Segundo o autor, supondo-se que há perfeita mobilidade de mão de obra entre as regiões, só pode haver diferencial de renda per capita entre as regiões se as características dos indivíduos das regiões diferirem. Dessa forma, as políticas de desenvolvimento baseadas em subsídios e acumulação de capital físico que foram adotadas no Brasil sempre tentaram focar o primeiro problema mas na verdade são mais adequadas ao segundo enfoque. Barros (1993), Barros e Mendonça (1995) e Barros, Camargo e Mendonça (1997) reforçam a tese da importância das características dos indivíduos, notadamente de educação, na determinação da diferença de renda através de evidências empíricas.

Bourguignon, Ferreira e Leite (2002) investigou as diferenças entre as distribuições de renda do Brasil,

EUA e México utilizando uma extensão de Oaxaca (1973) e de Blinder (1973), que consiste na simulação de distribuições contrafactuais construídas a partir da substituição dos valores originais dos parâmetros da distribuição pelos da distribuição de outro país. Dessa forma pode-se medir o efeito na distribuição de renda de um país caso alguma característica dos indivíduos, representada por um parâmetro da distribuição de renda, seja equiparada a de outro país. Concluiu-se que a desigualdade de dotação de capital humano e transferências explica cerca de 2/3 da diferença de desigualdade entre o Brasil e o EUA.

Neste trabalho analisaremos os efeitos de uma alteração nas características educacionais de uma região<sup>1</sup> sobre a sua distribuição de renda do trabalho. Utilizamos uma abordagem não-paramétrica apresentada em DiNardo, Fortin e Lemieux (1996) quando analisaram os efeitos de fatores institucionais e do mercado de trabalho, tais como valor do salário mínimo e grau de sindicalização, sobre a distribuição de salários-hora<sup>2</sup> nos Estados Unidos. Nessa metodologia os efeitos sobre o valor dos salários de uma determinada característica da população ou fator que influencie o comportamento da mesma são mensurados ao aplicar-se um método não-paramétrico de estimação da função de densidade, reponderando-se as amostras pelo característica/fator que se pretende analisar. Constroem-se dessa forma o que denominamos de distribuição contrafactual, a qual pode ser comparada com a distribuição original dos salários da população. No nosso caso, trabalharemos com a distribuição da renda do trabalho, uma vez que não se dispõe de dados de salários para o Brasil.

Essa abordagem tem o mesmo fundamento da decomposição de Oaxaca (1973)<sup>3</sup>, que se baseia na construção de contrafactuais, mas ao contrário deste - que realizou seu trabalho analisando somente as médias das distribuições - analisaremos toda a distribuição. A metodologia que empregamos nos permite visualizar claramente a função de densidade da renda do trabalho, assim como observar as alterações que ocorreriam nessa distribuição caso a população sofresse qualquer alteração nas condições educacionais, mantendo-se a estrutura de remuneração do trabalho originalmente vigente. Neste sentido, ao invés de focarmos o estudo de diferencial de renda e educação em algum índice de desigualdade, como por exemplo os coeficientes de Gini e de Theil ou a variância do logaritmo dos salários, pretendemos extrair nossas conclusões observando toda a distribuição de renda do trabalho, usando medidas de diferenças de distribuições e teste de desigualdade de distribuições de Kolmogorov-Smirnov.

A construção das distribuições contrafactuais se faz reponderando a amostra conforme alguma característica que se pretende estudar. Ou seja, para estudarmos, como no nosso caso, os efeitos da escolaridade sobre a renda do trabalho, estimamos a distribuição da renda do trabalho reponderando-se as amostras disponíveis de tal forma que as mesmas passem a compor uma população com o perfil de escolaridade que se deseja. Dessa forma, podemos obter a distribuição da renda do trabalho de uma região caso ele tivesse o perfil de escolaridade de outra região, mantendo a sua estrutura de remuneração original. Deve ficar claro que existe uma limitação dessa metodologia uma vez que são considerados apenas efeitos parciais, não sendo possível analisar os efeitos em equilíbrio geral. No entanto esta abordagem será bastante útil para responder

---

<sup>1</sup>Entende-se aqui como região uma determinada área geográfica da qual desejamos estudar uma determinada característica dos seus habitantes, podendo ser um município, estado, conjunto de estados, região geográfica ou país.

<sup>2</sup>A partir desse ponto a palavra salário será usada para se referir a salário-hora ou, em inglês, *wage*.

<sup>3</sup>A abordagem de Oaxaca é geralmente restrita à comparação de médias. Quando as distribuições são unimodais, simétricas e têm variâncias semelhantes este procedimento é bem qualificado para analisar as mudanças na estrutura dos salários. No entanto, não se pode esperar que estas condições sejam válidas quando estamos comparando distribuições de salários de regiões diferentes. Daí a opção por uma estimação de toda a função de densidade.

perguntas do tipo: qual seria a distribuição de renda do trabalho no Estado do Ceará se as condições de educação fossem semelhantes às do Estado de São Paulo?

No presente trabalho mensuramos a mudança na distribuição de renda do trabalho do Estado do Ceará e da Região Nordeste caso a população residente nessas regiões obtivessem a mesma escolaridade do Estado de São Paulo e da Região Sudeste<sup>4</sup>.

Na seção seguinte apresentamos alguns fatos estilizados que ilustram o problema de desigualdade no Brasil. Na seção 3 apresentamos os dados utilizados na análise a que se propõe este trabalho. Na seção 4 discutimos detalhadamente a metodologia de estimação de densidade de núcleo ponderada (*weighted kernel density estimator*) e a construção de densidades contrafactuais, além de apresentarmos algumas medidas paramétricas e não-paramétricas utilizadas para comparar as densidades estimadas. Os resultados são apresentados na seção 5, e finalmente, na seção 6, concluímos.

## 2 Fatos Estilizados

A desigualdade entre as regiões e estados brasileiros pode ser verificada tanto nos indicadores de bem-estar da população como no nível de renda dos seus habitantes. A diferença entre os Índices de Desenvolvimento Humano (IDH) das regiões diminuiu entre 1991 e 2000; a distância entre os IDHs das regiões Nordeste e Sudeste, por exemplo, que era 0,16 em 1991, passou para 0,12 em 2000 (Tabela 1)<sup>5</sup>. No entanto a posição relativa das regiões não mudou desde 1970, quando o índice foi computado pela primeira vez. Além disso, a posição relativa dos estados também não apresentou expressiva mudança no mesmo período, haja visto que os nove estados nordestinos sempre se situaram entre as 11 unidades da federação com piores IDH no período compreendido entre 1970 e 2000. A desigualdade de renda, medida pelos coeficientes de Theil e de Gini<sup>6</sup>, se agravou em todas as regiões brasileiras nas décadas de 70 e 80, e apresentou uma leve melhora na década seguinte; no entanto a piora foi mais acentuada nas Regiões Norte e Nordeste, o que as tornou ainda mais desiguais quando comparadas com a Região Sudeste (Tabela 2).

Tabela 1  
Índice de Desenvolvimento Humano - IDH

Índice de Desenvolvimento Humano	Ano	Brasil	Regiões				
			Centro-Oeste	Nordeste	Norte	Sudeste	Sul
IDH-Educação	1991	0.745	0.778	0.606	0.705	0.812	0.804
	2000	0.849	0.877	0.762	0.818	0.886	0.895
IDH-Longevidade	1991	0.668	0.682	0.587	0.637	0.709	0.715
	2000	0.731	0.747	0.669	0.706	0.759	0.776
IDH-Renda	1991	0.681	0.699	0.564	0.614	0.732	0.689
	2000	0.723	0.747	0.614	0.640	0.768	0.747
IDH-Total	1991	0.698	0.720	0.586	0.652	0.751	0.736
	2000	0.768	0.790	0.682	0.722	0.805	0.806

Fonte: autores

<sup>4</sup> Como será visto na seção seguinte, de fato não consideramos a Região Sudeste integralmente.

<sup>5</sup> Os IDHs das regiões foram calculado pelos autores, seguindo a metodologia do PNUD.

<sup>6</sup> O valor de ambos os coeficientes varia entre 0, quando não há desigualdade, e 1, quando a desigualdade é máxima.

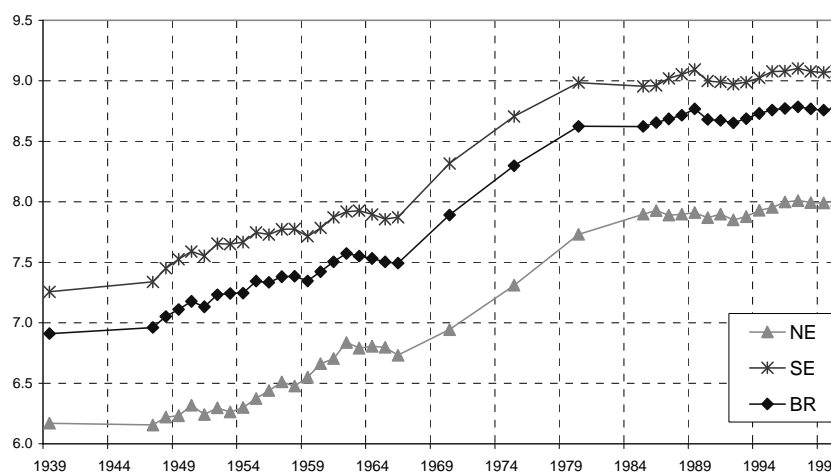
Tabela 2  
Coeficientes de Theil e de Gini da Renda do Trabalho

	Theil			Gini		
	1970	1980	1991	1981	1990	2001
Região Norte	0,44	0,56	0,72	0,51	0,58	0,57
Região Nordeste	0,57	0,65	0,78	0,57	0,63	0,60
Região Centro-Oeste	0,55	0,66	0,70	0,58	0,61	0,60
Região Sudeste	0,61	0,60	0,66	0,56	0,58	0,57
Região Sul	0,53	0,58	0,63	0,54	0,58	0,55
<b>Brasil</b>	<b>0,68</b>	<b>0,70</b>	<b>0,78</b>	<b>0,58</b>	<b>0,61</b>	<b>0,60</b>

Fonte: IPEA

Em termos de PIB *per capita*, as diferenças regionais se mantêm desde a década de 40 (gráfico 1); nem mesmo a participação relativa da Região Nordeste no PIB nacional se alterou, de fato, diminuiu de 16,7% em 1939 para 14,8% em 1960, e chegou a 13,1% em 2000<sup>7</sup>.

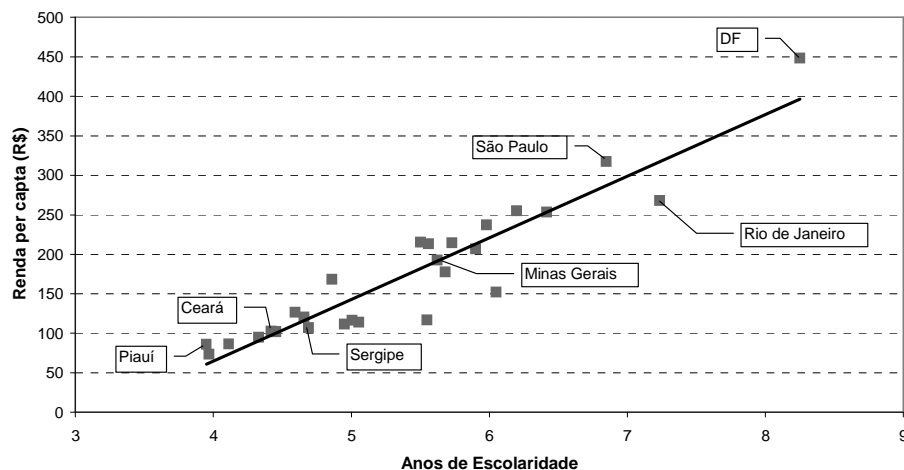
Gráfico 1  
Renda *per Capita* (em log)



Os dados do censo de 2000 mostram que há uma relação linear e positiva entre a renda *per capita* nos Estados e a escolaridade média da sua população acima de 25 anos (gráfico 2), o que corroboraria a tese de que as disparidade regionais de renda refletem as diferenças de capital humano de seus habitantes.

<sup>7</sup>Dados de PIB e população por região do IBGE.

Gráfico 2  
Renda *per Capita* x Escolaridade Média



### 3 Dados

Utilizamos os dados da Pesquisa Nacional por Amostras Domiciliares (PNAD) do ano de 1999 com as ponderações obtidas a partir dos resultados do censo de 2000. Os resultados da PNAD são apresentados tanto em termos de indivíduos como em termos de domicílios. No caso da pesquisa por indivíduos, que é a de nosso interesse, dispõe-se de informações sobre as características gerais (UF de residência atual, idade, gênero, raça, cor, etc), características de migração (UF onde nasceu, UF onde morava 5 anos antes da pesquisa, etc), educação, trabalho e renda dos pesquisados. A cada indivíduo entrevistado é atribuído um peso, que traduz o quanto as pessoas com as suas características representam em relação à população.

A amostra utilizada consistiu nos indivíduos com renda do trabalho positiva no mês de referência da pesquisa (setembro/1999), com escolaridade conhecida e com jornada de trabalho maior ou igual a 40 horas semanais. Excluiu-se as pessoas com jornada inferior a 40 horas a fim de se tentar uniformizar o tamanho da jornada e aproximar as medidas de renda do trabalho e de salário, uma vez que essa última variável seria a ideal para esse tipo de estudo.

Foram entrevistadas 352.393 pessoas em todo o território nacional, sendo que 67.111 tinham menos de 10 anos de idade a época, e para as quais não se aplica o questionário de trabalho e renda. Das 285.282 pessoas restantes, somente 173.634 eram economicamente ativas<sup>8</sup>, e destas, 3.265 pessoas tinham escolaridade ou renda ignorados, restando portanto 170.369 pessoas. A partir dessa amostra, que engloba indivíduos em todo o Brasil, extraímos 5 sub-amostras de acordo com a UF de residência do entrevistado: Estado do Ceará (CE), Estado de São Paulo (SP), Região Nordeste (NE), Região Sudeste excluindo o Estado do Espírito Santo (SE1), e Região Sudeste excluindo os Estados do Espírito Santo e São Paulo (SE2)<sup>9</sup>.

<sup>8</sup>Compreendem a população economicamente ativa as pessoas ocupadas ou que estão empreendendo esforços a procura de ocupação.

<sup>9</sup>SE2 compreende os Estados do Rio de Janeiro e Minas Gerais, respectivamente a segunda e terceira unidades mais ricas da

Escolheu-se as Regiões Nordeste e Sudeste por se tratarem das regiões com menor e maior renda per capita, respectivamente. Os Estados do Ceará e de São Paulo foram escolhidos entre os da Região Nordeste e da Região Sudeste, respectivamente, pelo fato do primeiro, mesmo sendo um dos estados mais pobres da Federação, ser apontado como exemplo de sucesso na implantação de políticas de desenvolvimento baseadas na atração de investimentos industriais privados, e o segundo por se tratar do ente mais rico da Federação. A exclusão do Estado do Espírito Santo se deve ao fato de sua economia ser bem menos desenvolvida e industrializada que a dos demais estados da região a ponto do mesmo ser também alvo das políticas públicas de combate à desigualdade regional, fazendo parte da área de atuação da SUDENE. Optou-se também por considerar uma sub-amostra que não englobasse o Estado de São Paulo, por se tratar de um estado peculiar no que tange à sua renda per capita e ao seu elevado grau de industrialização, que são muito superiores aos dos demais estados da Região Sudeste e da Federação. Ademais, como veremos adiante, São Paulo é o único estado no qual o valor do salário mínimo, que é determinado exogenamente pela União, não se constitui numa forte restrição à distribuição de renda do trabalho. Nos demais estados a moda da distribuição da renda do trabalho é exatamente o valor do salário mínimo.

As tabelas a seguir apresentam os quantitativos da amostra utilizada por região/estado de interesse e as principais estatísticas descritivas, que foram calculadas considerando-se a ponderação de cada indivíduo.

Tabela 3 - Quantitativo por Região/Estado

	Brasil	NE	CE	SE1	SE2	SP
Pessoas pesquisadas	352.393	113.902	22.124	110.558	67.890	42.668
PEA com escolaridade e renda conhecidos	170.369	53.210	10.163	53.162	32.658	20.504
Pessoas com renda do trabalho nula	37.808	14.718	2.588	9.530	6.061	3.469
Pessoas com jornada inferior a 40hr/semana	28.501	10.237	1.867	8.251	5.486	2.765
Tamanho da amostra	104.060	28.255	5.708	35.381	21.111	14.270

Tabela 4 - Estatísticas Descritivas para Amostra Seleccionada

(em R\$)	Brasil	NE	CE	SE1	SE2	SP
Renda do trabalho						
média	572	358	338	676	554	775
desvio-padrão	905	704	697	977	866	1.055
Coef. Gini	0,55	0,57	0,58	0,52	0,53	0,50
Log da renda do trabalho						
média	5,81	5,30	5,21	6,04	5,83	6,22
desvio-padrão	0,96	0,95	0,98	0,90	0,90	0,85
Escolaridade (anos)	6,4	4,6	4,5	7,3	6,9	7,8
Coef. Gini	0,32	0,42	0,42	0,29	0,30	0,27

---

Federação.

Verifica-se a partir da tabela 4 que quanto maior a renda média, menor a desigualdade de renda e de escolaridade, medidas aqui pelo coeficiente de variação de Pearson e pelo coeficiente de Gini. Isso nos permite levantar a hipótese que a desigualdade de renda e de escolaridade são fortemente correlacionadas, e por sua vez menores nas regiões/estados mais ricos. A escolaridade média nas regiões mais pobres é cerca de 3 anos menor que nas regiões mais ricas. De fato, observa-se que a renda é diretamente proporcional à escolaridade, o que vem a reforçar a hipótese de que o diferencial de renda pode ser explicada pela diferença de escolaridade.

Computamos para cada região/estado a mediana do logaritmo da renda por nível de escolaridade. As curvas escolaridades x renda, também conhecidas como curvas de Mincer, se encontram plotadas nos gráficos 3(a) e (b). Depreende-se que o retorno marginal da escolaridade é crescente para níveis de escolaridade acima de 10 anos, ao contrário do que se observa nos países desenvolvidos com alto nível escolaridade da sua população.

Gráfico 3(a) - Escolaridade x Renda

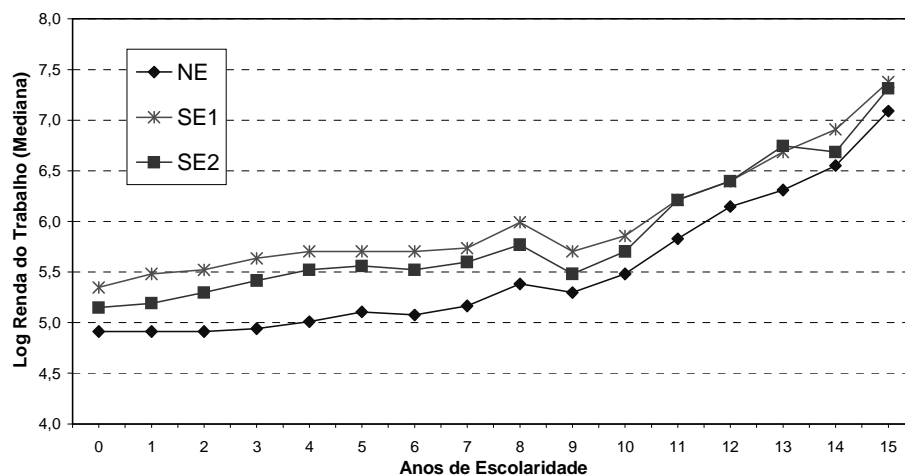
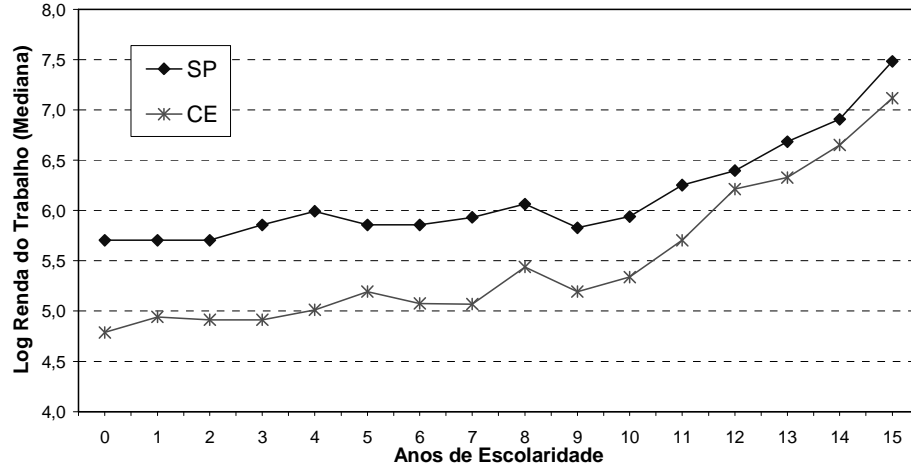


Gráfico 3(b) - Escolaridade x Renda



Apesar de dispormos dos dados de um expressivo número de pessoas entrevistadas, como pode ser observado na última linha da tabela 3, a quantidade de informação contida nessas entrevistas é bem menor do que aparenta. A renda, além de ter sido coletada como um valor inteiro, foi quase sempre reportada em múltiplos de R\$10 ou do salário mínimo (R\$136 à época) e ainda os valores concentram-se bastante em múltiplos de R\$100. A quantidade de níveis de renda do trabalho situam-se entre 480, para CE e 1040, para SE1.

## 4 Metodologia

Utilizaremos um modelo semiparamétrico para construir funções de densidade contrafactuais. Essas densidades contrafactuais são estimadas a partir de uma amostra que se gerou tomando por base a amostra original dos dados alterando o atributo do qual se deseja estudar o impacto sobre a distribuição de renda. O método compreende duas etapas: a primeira, paramétrica, que resume-se à construção de funções de reponderação; e a segunda, não paramétrica, que consiste na estimação, baseada em funções núcleo, de funções de densidade, conforme proposto por Rosenblatt (1956) e Parzen (1962).

Seja,  $\hat{f}_h$  a estimativa de densidade de núcleo da densidade  $f$ , cujo suporte é a variável  $w$ , baseada em uma amostra aleatória de tamanho  $n$ ,  $\{W_1, W_2, \dots, W_n\}$ , com ponderação  $\theta_1, \dots, \theta_n$ , respectivamente, e onde  $\sum_i \theta_i = 1$ . Temos então:

$$\hat{f}_h(w) = \sum_{i=1}^n \frac{\theta_i}{h} K\left(\frac{w - W_i}{h}\right) \quad (1)$$

onde  $h$  é a janela e  $K(\cdot)$  é a função núcleo. Os núcleos mais utilizados são o uniforme, o gaussiano e o de Epanechnikov, sendo que a sua escolha é uma decisão ad-hoc do econometrista, que deve levar em conta a natureza da variável cuja densidade está sendo estimada. No presente trabalho, acompanhando as

sugestões de DiNardo, Fortin e Lemieux (1996) e Butcher e DiNardo (1998), adotamos o núcleo gaussiano e trabalhamos com o logaritmo da renda do trabalho para reduzir o problema de assimetria.

A escolha da janela é um ponto importante em estimações de densidade de núcleo, pois que há um *trade-off* entre viés (diferença entre a distribuição estimada e a real) e variância: janelas maiores resultam em maior viés e menor variância, e vice-versa. Há diversos métodos de seleção automática da janela, entre os quais destacam-se os métodos de validação cruzada e os métodos de *plug-in*<sup>10</sup>. No entanto esses métodos não são adequadas para dados com as características dos que dispomos, visto que são censurados por intervalos (agrupados). Os métodos de validação cruzada, por exemplo, como apontado por Silverman (1986), tendem a gerar resultados inadequados,  $h = 0$ . Adotamos então o método visual de seleção, conforme procedimento detalhado em Butcher e DiNardo (1998): iniciamos com uma janela extremamente estreita (baixa suavização),  $h = 0,05$ , e aumentamo-la até obtermos uma distribuição suave, o que veio a acontecer com  $h = 0,12$ . Justifica-se o procedimento de se iniciar com uma janela pequena e ir aumentando-a pela crença de que o olho humano é mais habilidoso em suavizar do que no sentido contrário (Butcher e DiNardo (1998)). Subjacente a esse método estamos adotando a hipótese que a distribuição de renda do trabalho, e conseqüentemente a produtividade marginal do trabalho, é suave, o que nos parece bastante razoável haja vista o tamanho da população. A escolha do limite inferior das janelas que produzem distribuições suaves indica que priorizamos o viés em detrimento da variância.

A estimação de densidades contrafactuais é realizada conforme proposto por DiNardo, Fortin e Lemieux (1996), onde se escolhe funções de reponderação da amostra. Pode-se considerar que cada observação da amostra é um vetor  $(w, z)$ , onde  $w$  representa os salários (uma variável contínua),  $z$  os atributos de cada indivíduo (consideraremos apenas a variável educação, medida em anos de escolaridade). Assim temos distribuições de densidade conjuntas  $F(w, z)$  de renda do trabalho e educação. Assim, a densidade da renda do trabalho da região 1,  $f_{R1}(w)$ , pode ser escrita como a integral da densidade de renda do trabalho condicionada no nível de educação dos indivíduos,  $f(w/z)$ , sobre a distribuição de educação,  $F(z)$ :

$$\begin{aligned} f_{R1}(w) &= \int_{z \in \Omega_z} dF(w, z | E_{w,z} = R1) \\ &= \int_{z \in \Omega_z} f(w|z, E_w = R1) dF(z | E_z = R1) \\ &\equiv f(w; E_w = R1, E_z = R1) \end{aligned} \tag{2}$$

onde  $\Omega_z$  é o domínio do conjunto de atributos,  $E_z$  representa a região de onde se considera a distribuição de educação e  $E_w$  representa a região de onde se considera a distribuição de renda do trabalho. Para fazer o estudo de contrafactuais estaremos interessados em modificar a estrutura de atributos e portanto definiremos como  $f(w; E_w = R1, E_z = R1)$  a densidade real de renda do trabalho da região 1 e  $f(w; E_w = R1, E_z = R2)$  a densidade de renda do trabalho da região 1 que prevaleceria se a distribuição de educação fosse aquela existente na região 2 no mesmo período.

Sob a hipótese de que a estrutura de renda do trabalho da região 1,  $f(w|z, E_w = R1)$ , não depende da distribuição de educação na região 2, podemos escrever a densidade hipotética  $f(w; E_w = R1, E_z = R2)$

<sup>10</sup>Ver Park e Marron (1990) e Sheater e Jones (1991) como exemplos.

como,

$$\begin{aligned} f(w; E_w = R1, E_z = R2) &= \int f(w|z, E_w = R1) dF(z|E_z = R2) \\ &\equiv \int f(w|z, E_w = R1) \Psi_z(z) dF(z|E_z = R1) \end{aligned} \quad (3)$$

onde  $\Psi_z(z)$  é uma função de reponderação definida por

$$\Psi_z(z) \equiv dF(z|E_z = R2)/dF(z|E_z = R1) \quad (4)$$

A equação (3) define a densidade de renda do trabalho na região 1 que prevaleceria se as condições educacionais fossem similares às da região 2 e conforme pode ser observado é idêntica à definição em (2), exceto pela função de reponderação  $\Psi_z(z)$ . Na verdade, o problema de estimação da função de densidade contrafactual desejada fica reduzido ao cálculo de ponderações apropriadas. Logo, poderemos estimar funções de densidade contrafactuais usando o método de estimadores de núcleo ponderados onde usamos um novo ponderador que contém uma estimativa para  $\Psi_z(z)$ . Assim temos,

$$\hat{f}(w; E_w = R1, E_z = R2) = \sum_{i \in S_{R1}} \frac{\theta_i}{h} \hat{\Psi}_z(z_i) K\left(\frac{w - W_i}{h}\right) \quad (5)$$

onde  $\sum_{i \in S_{R1}} \theta_i \hat{\Psi}_z(z_i) = 1$  e  $S_{R1}$  é o conjunto de índices da amostra de indivíduos da região 1. As diferenças observadas entre a densidade real da região 1 e a densidade hipotética criada representa o efeito de uma mudança na distribuição de educação.

Pode-se observar que aplicando uma regra de Bayes em (4) este quociente pode ser escrito por

$$\Psi_z(z) = \frac{\Pr(E_z = R2|z) \Pr(E_z = R1)}{\Pr(E_z = R1|z) \Pr(E_z = R2)}. \quad (6)$$

Uma vez que o nível de educação é uma variável discreta que assume um número finito de valores, a estimação de  $\Psi_z(z)$  por um modelo probit é equivalente a uma simples contagem.

Há diversas formas, paramétricas e não-paramétricas, de comparar as densidades estimadas (reais e contrafactuais) e medir a diferença entre as mesmas. Pode-se simplesmente tomar a diferença entre as densidades contrafactual e real, obtendo-se uma completa descrição das mudanças na distribuição da renda do trabalho oriundas da reponderação pela escolaridade. Outras medidas de comparação entre as densidades estimadas tentam reduzir em um único número as diferenças entre as mesmas: distância de Kullbach-Leibler, de Sibson, de Chernoff, diferença entre os desvios-padrões, diferenças entre os percentis, diferenças entre os diferenciais de percentis (10-90, 10-50, 25-75, 5-95). Todas essas medidas são utilizadas neste trabalho. Além disso, faremos o teste de desigualdade de densidades de Kolmogorov-Smirnov.

## 5 Resultados

Aplicamos a metodologia descrita na seção anterior e estimamos as distribuições reais do logaritmo da renda do trabalho para as 5 sub-amostras e para a amostra completa (Brasil), conforme apresentado na seção

3. Em seguida, estimamos as distribuições contrafactuais para o Nordeste e para o Ceará reponderando as amostras pelas características de escolaridade da Região Sudeste (SE1 e SE2) e do Estado de São Paulo. Relatamos e comentamos nesta seção os resultados para o Estado do Ceará, reponderado pela escolaridade de São Paulo, e para a Região Nordeste, reponderada pela escolaridade da Região Sudeste (SE1 e SE2). Os resultados para os demais casos estão apresentados no apêndice.

Em cada um dos gráficos abaixo, 4 a 6, apresentamos três densidades: uma densidade contrafactual, a densidade real da qual essa se originou e a densidade real da região usada na reponderação<sup>11</sup>. O eixo horizontal dos gráficos está em escala logarítmica.

Gráfico 4

Densidades reais para SE1 e NE e contrafactual para NE com escolaridade de SE1

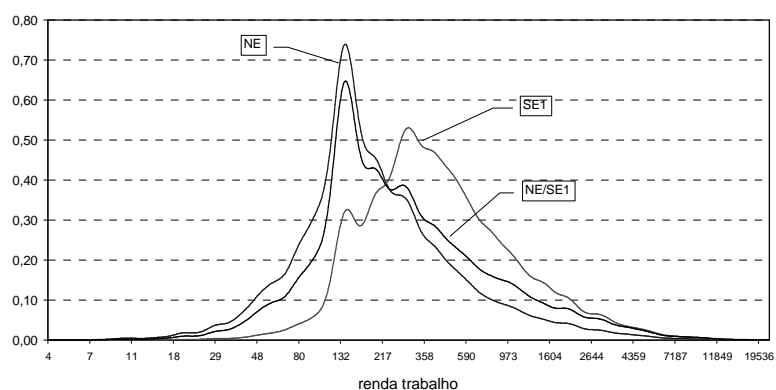
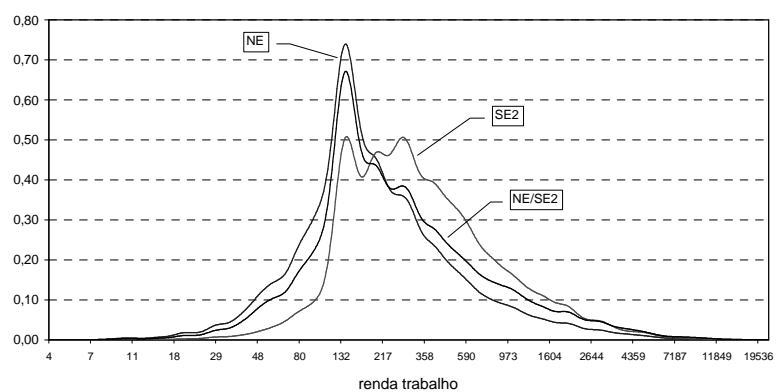


Gráfico 5

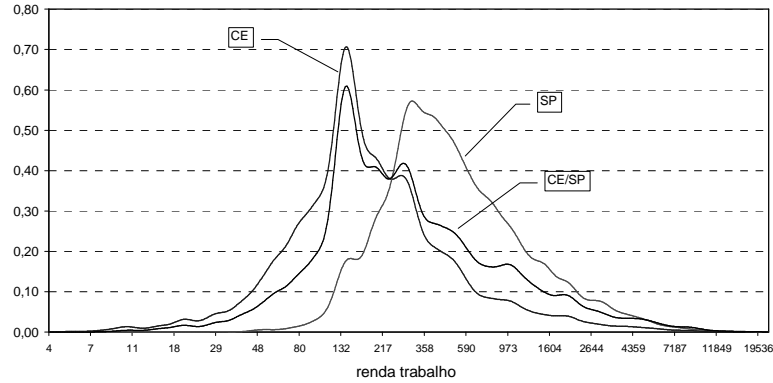
Densidades reais para SE2 e NE e contrafactual para NE com escolaridade de SE2



<sup>11</sup> Todos as grandezas monetárias estão expressas em Reais.

Gráfico 6

Densidades reais para SP e CE e contrafactual para CE com escolaridade de SP



Testamos a igualdade entre as distribuições empíricas originais e as suas contrafactuais usando o teste de Kolmogorov-Smirnov, no qual a estatística de teste é o máximo da diferença entre as densidades acumuladas:  $\max |F_1(w) - F_2(w)|$ . Rejeitamos a hipótese nula de que as densidades são iguais a 1% de significância em todos os casos, evidenciando que a mudança do perfil de escolaridade altera a distribuição de renda do trabalho.

Fica evidente que as distribuições <sup>12</sup> são sensíveis aos diferenciais de escolaridade e que o grau de sensibilidade depende do percentil de renda considerado. O valor do salário mínimo, R\$136 à época da pesquisa, é fator determinante na renda dos trabalhadores do Nordeste, do Ceará e do Sudeste (SE2), haja visto a enorme concentração de massa em torno desse valor na distribuição. A distribuição de renda da Região Sudeste (SE2) é bimodal, sendo que uma das modas corresponde ao valor do salário mínimo e a outra é igual à moda da distribuição de São Paulo, cerca de R\$300. As distribuições de SP e SE1 são bem parecidas o que mostra a significância de São Paulo sobre o agregado da Região Sudeste. Pode-se afirmar então que em todos os estados das Regiões Nordeste e Sudeste, com exceção de São Paulo (SE2), o valor do salário mínimo tem forte impacto sobre a distribuição de renda do trabalho.

Comparando-se as caudas das distribuições contrafactuais e das distribuições originais verifica-se que as mesmas se aproximam bastante a menos de uma translação à direita das primeiras. Isso é decorrente da proporcionalidade direta entre escolaridade e renda em conjunto com a reponderação da amostra, cujos efeitos sobre a distribuição de renda assemelham-se ao que seria obtido caso adicionássemos uma constante à escolaridade de cada indivíduo. Observa-se que as áreas centrais das distribuições de renda, que é exatamente onde se concentram mais indivíduos e aqueles que têm renda próxima a um salário mínimo, pouco se modificam após a reponderação. Justifica-se esse comportamento pelo fato da estrutura salarial não ser alterada nesse processo, sobrepondo-se ao efeito do aumento de escolaridade resultante do mesmo processo. A moda, por exemplo, permanece igual a um salário mínimo em todos os casos .

<sup>12</sup>Leia-se nos gráficos:

NE/SE1: contrafactual do Nordeste usando escolaridade do Sudeste (MG+RJ+SP).

NE/SE2: contrafactual do Nordeste usando escolaridade do Sudeste (MG+RJ).

CE/SP: contrafactual do Ceará usando escolaridade de S.Paulo.

A tabela 5 apresenta alguns pontos notáveis (percentis, médias, desvios-padrão e diferenças de percentis) das distribuições estimadas, reais e contrafactuais. O suporte das distribuições estimadas é o logaritmo da renda do trabalho, o qual delimitamos para efeitos computacionais ao intervalo [1,10] com passo de 0,01. A partir dessas distribuições construímos distribuições com o suporte sendo o nível da renda do trabalho, simplesmente tomando o exponencial de cada ponto das distribuições estimadas e renormalizando. Apresentamos nas tabelas A.1 e A.2 (no apêndice) os pontos notáveis de todas as distribuições construídas e as razões dos percentis das distribuições estimadas.

Tabela 5 - Pontos Notáveis das Distribuições de Renda do Trabalho Estimadas

Percentil	Distribuições Estimadas(*)								
	Brasil	CE	SP	NE	SE1	SE2	CE/SP	NE/SE1	NE/SE2
10	116	61	187	70	144	125	87	89	84
20	147	90	250	101	196	153	125	124	119
30	189	117	302	124	250	192	147	144	140
40	237	136	358	141	299	235	183	174	164
50	293	158	428	166	365	284	233	219	204
60	365	194	523	202	450	358	296	281	262
70	469	247	672	262	572	459	412	384	351
80	665	334	916	365	796	639	639	578	523
90	1.130	567	1.510	639	1.339	1.086	1.200	1.086	982
Média	572	338	775	358	676	553	584	534	493
DP	905	697	1.056	704	977	855	1.091	687	929
Diferença de Percentis									
10-90	1.014	506	1.323	569	1.195	961	1.113	997	898
10-50	177	97	242	96	221	159	145	130	120
50-90	837	409	1.082	473	974	801	967	867	778
20-80	518	244	666	264	600	486	514	454	404

(\*) Os valores apresentados foram extraídos das distribuições estimadas (em log) tomando-se a exponencial de cada ponto.

As médias e os desvios-padrão foram calculados diretamente da amostra reponderada com a renda em nível.

Tabela 6 - Relação da Renda nos Percentis

Percentil	CE x SP	CE/SP x SP	NE x SE1	NE x SE2	NE/SE1 x SE1	NE/SE2 x SE2
10	33%	47%	49%	56%	62%	67%
20	36%	50%	52%	66%	63%	78%
30	39%	49%	50%	64%	58%	73%
40	38%	51%	47%	60%	58%	70%
50	37%	54%	45%	58%	60%	72%
60	37%	57%	45%	57%	63%	73%
70	37%	61%	46%	57%	67%	76%
80	36%	70%	46%	57%	73%	82%
90	38%	79%	48%	59%	81%	90%

Comparando-se as distribuições originais do Ceará e de São Paulo verifica-se que a renda da primeira é de aproximadamente 1/3 da segunda em todos os percentis (tabela 6). Por outro lado, ao compararmos a distribuição do Ceará reponderada pela educação de São Paulo com a distribuição original de São Paulo, observamos dois efeitos: (i) houve um ganho na distribuição de renda do Ceará em todos os percentis,

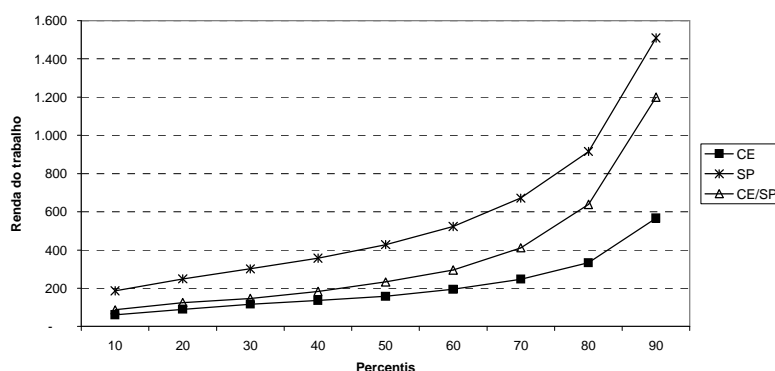
e (ii) o ganho foi maior quanto maior era o nível de renda (percentil)<sup>13</sup>. O retorno marginal crescente da escolaridade, ilustrado no gráfico 3, justifica o segundo efeito, pois os indivíduos mais escolarizados, e conseqüentemente com maior renda, ao receberem um “ganho” de escolaridade por efeito da reponderação terão suas renda elevadas em proporção maior que os menos escolarizados, cuja renda evolui pouco após um aumento de escolaridade.

Comparando-se a Região Nordeste com a Sudeste verificou-se que a renda da primeira se situa entre 45% e 50% da renda do Sudeste para a amostra SE1, e entre 55% e 65% considerando a amostra SE2. Após a reponderação, observaram-se os mesmos efeitos verificados quando comparamos a distribuição do Ceará com a de São Paulo. Observou-se que o efeito da reponderação é diretamente proporcional à diferença entre as distribuições originais, logo sendo maior no caso CE x SP e menor para NE x SE2, para todos os percentis. Uma comparação interessante é que a renda média do Nordeste reponderada pela educação da Região Sudeste (SE1) é 93% da renda média brasileira. Aqui, à semelhança da comparação CE x SP, o fato da escolaridade apresentar retorno marginal crescente justifica um ganho de renda maior nos decis mais altos.

Os ganhos absolutos são significativos em todos os caso: em nenhum decil o ganho de renda com a reponderação é inferior a 13% da renda original, chegando a atingir 70% no nono decil para a comparação CE x SP. O mais relevante é que a renda dos nordestinos nunca é inferior a cerca de 60% da renda do sudestino, para a amostra SE1 e a 67%, para a amostra SE2, e que nos decis superiores ela se aproxima bastante da renda dos moradores da região sudeste, se aqueles possuísem o mesmo nível de educação destes. A propósito, a renda do nono decil do NE reponderado pela escolaridade da região SE2 atinge 90% da renda de SE2.

Este fato está melhor representado nos gráficos 7 a 9 (e A.4 a A.6 no apêndice), onde a distribuição contrafactual e a distribuição de renda da região/estado que foi utilizado como base para a reponderação se aproximam nos níveis mais elevados de renda.

Gráfico 7 -Evolução da Renda do Trabalho por Percentis para CE, SP e CE/SP



<sup>13</sup>Para o primeiro decil a renda do Ceará evoluiu de 33% para 47% da renda de S. Paulo, e para o nono decil passou de 38% para 79%.

Gráfico 8 -Evolução da Renda do Trabalho por Percentis para NE, SE1 e NE/SE1

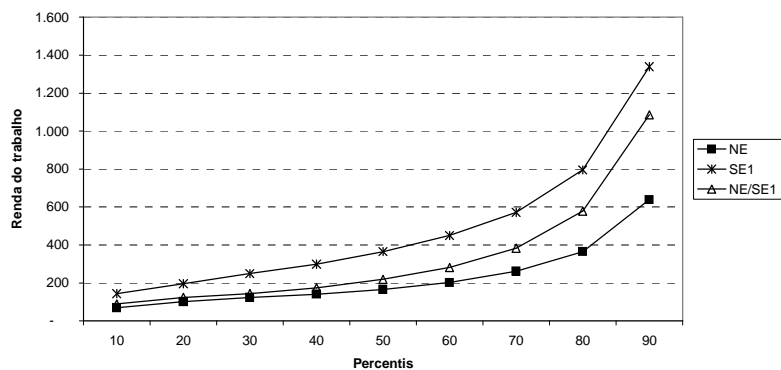
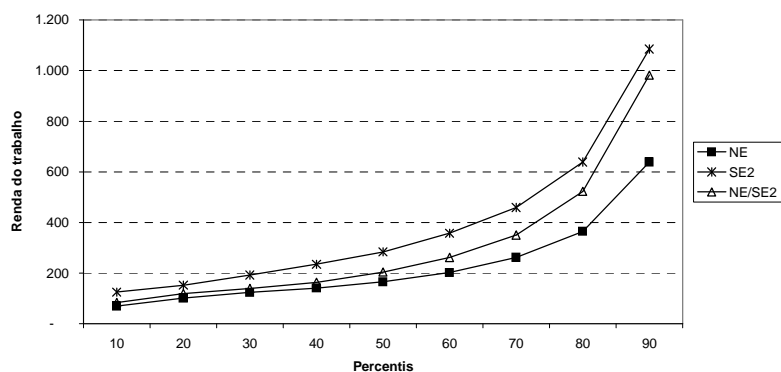


Gráfico 9 -Evolução da Renda do Trabalho por Percentis para NE, SE2 e NE/SE2



Nos gráficos 10 a 12 abaixo apresentamos a diferença entre as duas distribuições reais e a diferença entre a distribuição contrafactual e a real que a originou<sup>14</sup>. Esses gráficos nos permitem visualizar o quanto as distribuições de renda se aproximaram após a reponderação pela escolaridade.

<sup>14</sup>Para cada nível de renda tomou-se a diferença (contrafactual - real), ou seja,  $f(w; E_w = R1, E_Z = R2) - f(w; E_w = R1, E_Z = R1)$ .

Gráfico 10 - Diferença das Distribuições (SE1 x NE)

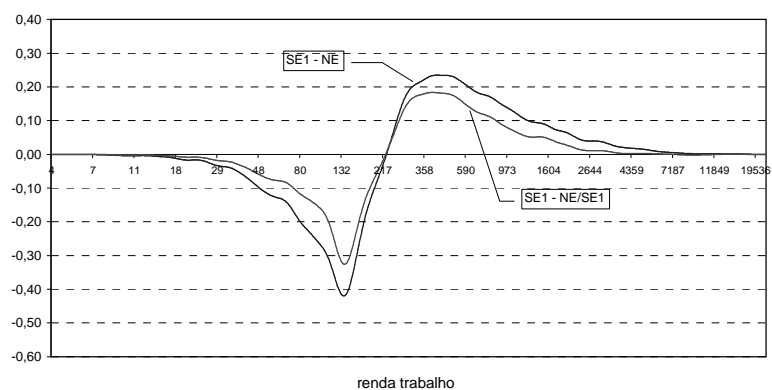


Gráfico 11 - Diferença das Distribuições (SE2 x NE)

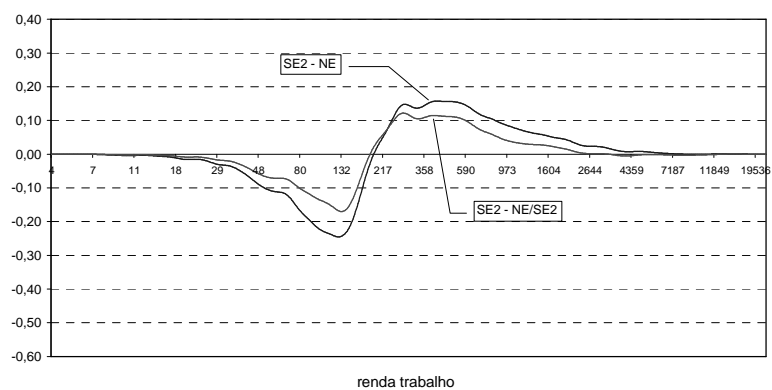
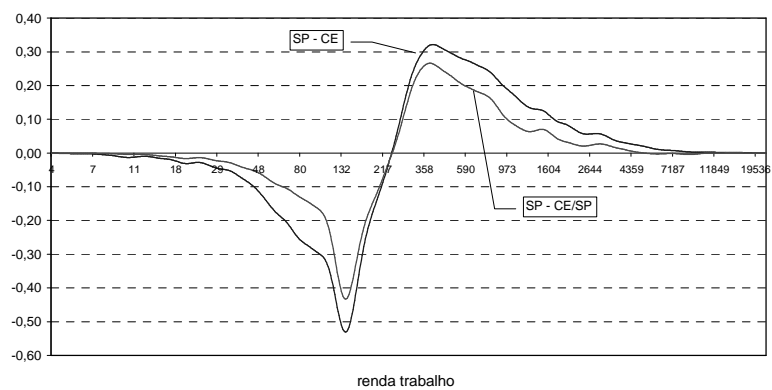


Gráfico 12 - Diferença das Distribuições (SP x CE)



Verificamos que as diferenças são negativas nos níveis inferiores de renda e positivas para níveis elevados de renda, o que denota que prepondera o efeito de diferença de média entre as distribuições em relação ao

efeito de diferença de dispersão. Caso esse último efeito prevalecesse observaríamos diferenças positivas nas caudas e negativas no centro do suporte das distribuições.

Computamos a distância entre as distribuições, antes e depois da reponderação, pela medida paramétrica conhecida como distância de Kullback-Leibler ( $J$ )<sup>15</sup>. Os resultados estão reportados na tabela 7 abaixo. Basicamente o que essa medida faz é representar a área entre duas as distribuições, onde cada elemento de área (correspondendo a cada  $w$ ) é ponderado pela diferença de densidade entre as distribuições. O efeito da ponderação é introduzir uma relação não-linear entre a distância das distribuições, para cada  $w$ , e sua contribuição para  $J$ . Verifica-se que mais de 55% das distâncias entre o Ceará e São Paulo e entre as regiões Nordeste e Sudeste (SE1 e SE2) são explicadas pela escolaridade - 55,3% da distância entre São Paulo e Ceará e 55,0% (57,3%) da distância entre o Nordeste e a região SE1 (SE2). Isto mostra que mais da metade da diferença de renda entre a região/estado mais pobre em relação ao mais rico é devido aos diferenciais de escolaridade da população. Computamos também a distância das distribuições pelas métricas de Chernoff<sup>16</sup> e Sibson<sup>17</sup>, sugeridas em Krzanowski (2003), e obtivemos resultados semelhantes, os quais se encontram apresentados na Tabela 9.

Tabela 7 - Distâncias de Kullback-Leibler

Distância entre as Distribuições								
	CE	NE	CE/SP	CE/SE1	CE/SE2	NE/SP	NE/SE1	NE/SE2
SP	1,5901	1,3826	0,7110			0,6700		
SE1	0,9608	0,8031		0,3969			0,3611	
SE2	0,5396	0,4127			0,2145			0,1761
% da distância entre as distribuições explicado pela educação								
	SP - CE	55,3%		SP - NE	51,5%			
	SE1 - CE	58,7%		SE1 - NE	55,0%			
	SE2 - CE	60,3%		SE2 - NE	57,3%			

Calculamos os coeficientes de entropia de Theil e de concentração de Gini das distribuições estimadas. Esses coeficientes são medidas de dispersão e são amplamente usadas como medidas de desigualdade quando aplicadas a distribuições de renda. Os resultados estão apresentados na tabela 8 abaixo (e na tabela A.4 no apêndice). Em suma, observou-se um aumento da desigualdade de renda quando reponderamos a amostra. Isto se deve ao fato de que a elevação de renda não é uniforme ao longo de toda a distribuição, observando maiores ganhos para os níveis maiores de renda.

<sup>15</sup>A distância de Kullback-Leibler,  $J$ , é uma medida de divergência entre duas distribuições  $f_1$  e  $f_2$  :

$$J = \int_0^\infty [f_1(w) - f_2(w)] \ln \left( \frac{f_1(w)}{f_2(w)} \right) dw$$

<sup>16</sup>A distância de Chernoff entre as distribuições  $f_1$  e  $f_2$  é definida como  $-\log \left( \int \sqrt{f_1(x) f_2(x)} dx \right)$ .

<sup>17</sup>A distância de Sibson entre as distribuições  $f_1$  e  $f_2$  é definida como  $-\frac{1}{2} \int f_1(x) \ln \frac{f_1(x)}{f(x)} + f_2(x) \ln \frac{f_2(x)}{f(x)}$ , onde  $f(x) \equiv \frac{1}{2} [f_1(x) + f_2(x)]$ .

Tabela 8 - Coeficientes de Gini e Theil

		Coeficiente de Gini	Coeficiente de Theil
Brasil		0,551	0,607
Regiões	CE	0,582	0,768
	NE	0,574	0,731
	SE1	0,523	0,539
	SE2	0,534	0,573
	SP	0,505	0,498
	CE/SP	0,622	0,797
	NE/SE1	0,608	0,766
	NE/SE2	0,604	0,768

Obs.: SE1 = MG + RJ + SP

SE2 = MG + RJ

Mensuramos o efeito da reponderação pela escolaridade computando a variação percentual de algumas métricas, calculadas a partir das distribuições estimadas. Os resultados estão reportados na Tabela 9 abaixo. Em termos de renda média ou considerando a distância de Kullbach-Leibler, a escolaridade foi capaz de explicar mais de 55% da diferença de renda, sendo que no caso da reponderação do Nordeste pela escolaridade da região SE2 esse índice atinge cerca de 70% na primeira métrica. A reponderação ocasionou dois efeitos: (i) as medidas de tendência central se aproximaram e, (ii) as medidas de dispersão se elevaram, sinalizando que a desigualdade de renda nas regiões mais pobres aumentariam caso herdasse o perfil de escolaridade das mais ricas, mantendo-se a estrutura de salários. Em outras palavras, o Nordeste com o perfil de escolaridade do Sudeste seria mais rico, porém mais desigual. O efeito a que nos referimos no item (ii) decorre, em parte, do fato de nosso exercício não levar em conta a mudança na estrutura salarial decorrente da mudança de escolaridade, ou seja, em equilíbrio geral esse fenômeno não seria observado em parte ou totalmente.

Tabela 9 - Percentual Explicado pela Escolaridade

Métrica	SP x CE	SE1 x NE	SE2 x NE
Média	56,3%	55,3%	69,2%
Mediana	27,8%	26,9%	32,6%
Coeficiente de Theil	-10,7%	-18,0%	-23,4%
Coeficiente de Gini	-51,9%	-66,9%	-75,0%
Distância de Kullbach-Leibler	55,3%	55,0%	57,3%
Distância de Chernoff	56,0%	55,5%	57,5%
Distância de Sibson	52,5%	53,6%	56,5%

A robustez dos resultados apresentados ao tamanho da janela foi investigado reestimando-se todas as distribuições com  $h = 0,09$  (-25% de  $h$  original) e  $h = 0,15$  (+25%), e comparando os resultados com os obtidos com o valor original de  $h$ . Verificou-se que as alterações nos resultados não os comprometem, tendo

ainda não se observado uma relação direta entre a variação do tamanho da janela e a variação dos indicadores apresentados nas tabelas 5 a 9.

Cabe ressaltar que a parcela da diferença de renda explicada pela diferença de escolaridade talvez seja maior que a apresentada nesse trabalho, uma vez que o preço da cesta de consumo dos indivíduos é diferente entre as regiões. Os serviços e bens não-comercializáveis são mais baratos nas regiões mais pobres, refletindo exatamente a diferença de renda do trabalho entre as regiões, vez que eles tendem a ser mais intensivos em mão de obra. O efeito da correção pelo poder de compra da renda em cada região seria menor nas camadas de renda mais baixa (consomem menos serviços) e maior para os indivíduos com renda elevada (consomem mais serviços), acentuando o efeito apresentado nos gráficos 7 a 9.

## 6 Conclusão

Neste trabalho procuramos identificar o quanto do diferencial de renda entre as regiões Nordeste e Sudeste e entre o estado do Ceará e São Paulo é explicado pelo diferencial de escolaridade da população. Usou-se um modelo semiparamétrico para construir funções de densidade contrafactuais, reponderando os indivíduos da região/estado base pela distribuição de educação da região a ser comparada. Estimamos as distribuições de renda do trabalho reais e contrafactuais do Estado do Ceará e da Região Nordeste reponderadas pelas escolaridades da Região Sudeste e do Estado de São Paulo.

Verificou-se que (i) a dispersão de renda é maior nas distribuições com menor média, ou seja, a desigualdade renda, que é enorme em todas as regiões, é maior nas regiões mais pobres; (ii) mais de 55% da diferença de renda do trabalho entre a Região Nordeste e a Região Sudeste e entre os Estados de São Paulo e Ceará, quando medidas pela distância de Kullback-Leibler ou em termos de renda média, se devem à diferença de escolaridade; (iii) a reponderação pela escolaridade aumentou em cerca de 50% a renda média da Região Nordeste e em mais de 70% a do Estado do Ceará; (iv) a renda do Nordeste reponderada pela escolaridade do Sudeste equivale a 93% da renda média brasileira; (v) quanto mais elevado for o percentil de renda considerado, maior é a contribuição da diferença de escolaridade para a diferença de renda e, (vi) a dispersão de renda das regiões mais pobres aumenta quando fornecemos a elas o nível de escolaridade das regiões mais ricas, mantendo-se o perfil salarial da região.

Vários são os fatores que podem estar determinando a diferença de renda não explicada pelo diferencial de escolaridade, entre os quais podemos citar a expectativa de vida dos habitantes, fatores étnicos, estrutura etária da população, qualidade da infra-estrutura existente, presença/ausência de estímulos ao desenvolvimento e fatores históricos. Uma extensão natural deste trabalho seria, aplicando a metodologia de contrafactuais, decompor o diferencial de renda em alguns destes fatores, além da escolaridade, de forma a conseguir explicar uma parcela maior do diferencial de renda. A importância relativa dos fatores posta em conjunto com o custo de eliminá-los pode ser de grande valor na orientação de políticas públicas de combate à “desigualdade regional”. Outras extensões interessantes seriam: (i) reponderarmos a distribuição de renda da Região Nordeste pela escolaridade do Brasil sem a Região Norte<sup>18</sup> e; (ii) reponderarmos a distribuição de renda de todas as regiões pelo perfil de escolaridade do Brasil, permitindo que se determine o tamanho das

---

<sup>18</sup>Isto porque a PNAD não entrevista a população rural, que é bastante significativa na Região Norte.

desigualdades regionais controlando-se pela escolaridade.

## References

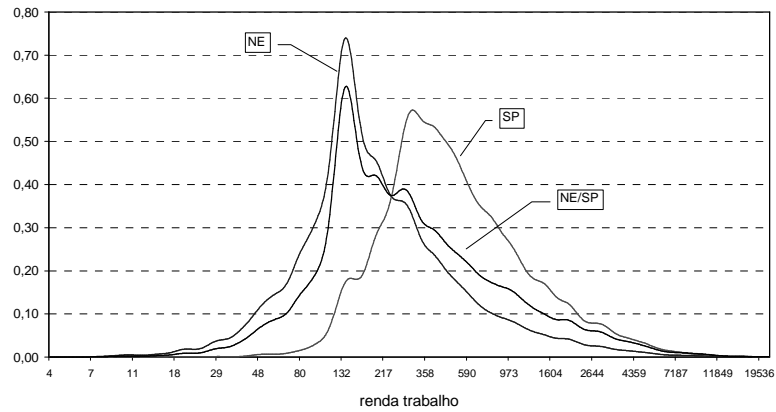
- [1] Barros, Ricardo Paes de (1993). “Regional Disparities in Education Within Brazil: the Role of Quality of Education”, Texto para Discussão n° 311, IPEA.
- [2] Barros, Ricardo Paes de & Mendonça, Rosane (1995). “Os Determinantes da Desigualdade no Brasil”, Texto para Discussão n° 377, IPEA.
- [3] Barros, Ricardo Paes de, Camargo, José Márcio & Mendonça, Rosane (1997). “A Estrutura do Desemprego no Brasil”, Texto para Discussão n° 478, IPEA.
- [4] Blinder, Alan S. (1973). “Wage Discrimination: Reduced Form and Structural Estimates”, *Journal of Human Resources*, 8, 436-455.
- [5] Bourguignon, François, Ferreira, Francisco & Leite, Phillippe (2002). “Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions Across Countries”, mimeo.
- [6] DiNardo, John, Fortin, Nicole M. & Lemieux, Thomas (1996). “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach”, *Econometrica*, Vol. 64, No. 5, 1001-1044.
- [7] Butcher, Kristin F. & DiNardo, John (1998). “The Immigrant and Native-Born Wage Distributions: Evidence from United States Censuses”, *NBER Working Paper Series 6630*.
- [8] Krzanowski, W. J. (2003). “Non-parametric estimation of distance between groups”, *Journal of Applied Statistics*, Vol. 30, No. 7, 743-750.
- [9] Oaxaca, R. (1973). “Male-Female Wage Differentials in Urban Labor Markets”, *International Economic Review*, 14, 693-709.
- [10] Park, B. U. & Marron, J. S. (1990). “Comparision of Data-driven Bandwidth Selectors”, *Journal of American Statistical Association*, 85, 66-72.
- [11] Parzen, E. (1962). “On Estimation of a Probability Density Function and Mode”, *The Annals of Mathematical Statistics*, 33, 1065-1076.
- [12] Pessoa, Samuel (2000). “Existe um Problema de Desigualdade Regional no Brasil?”, mimeo.
- [13] Rosenblatt, M. (1956). “Remarks on Some Non-parametric Estimates of a Density Function”, *The Annals of Mathematical Statistics*, 27, 832-837.
- [14] Sheater, S. J. & Jones, M. C. (1991). “A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation”, *Journal of Royal Statistical Society*, 53, 683-690.

[15] Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall, 1986.

## A Tabelas e Figuras

Gráfico A.1

Densidades reais para SP e NE e contrafactual para NE com escolaridade de SP



Diferenças entre as distribuições de SP e NE (real e contrafactual)

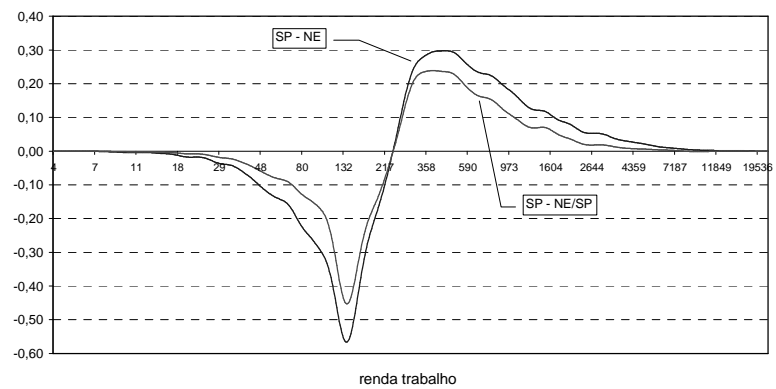
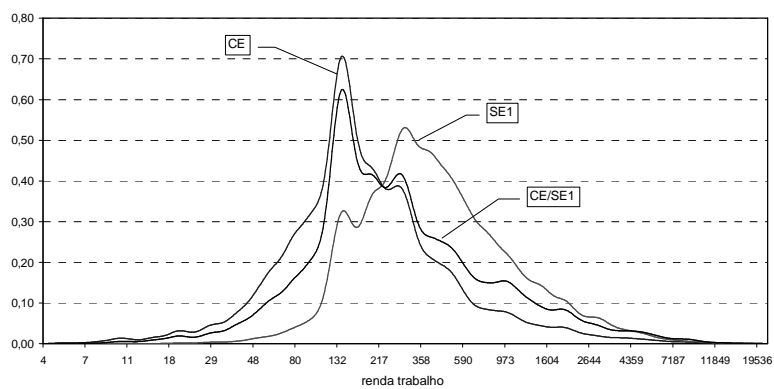


Gráfico A.2

Densidades reais para SE1 e CE e contrafactual para CE com escolaridade de SE1



Diferenças entre as distribuições de SE1 e CE (real e contrafactual)

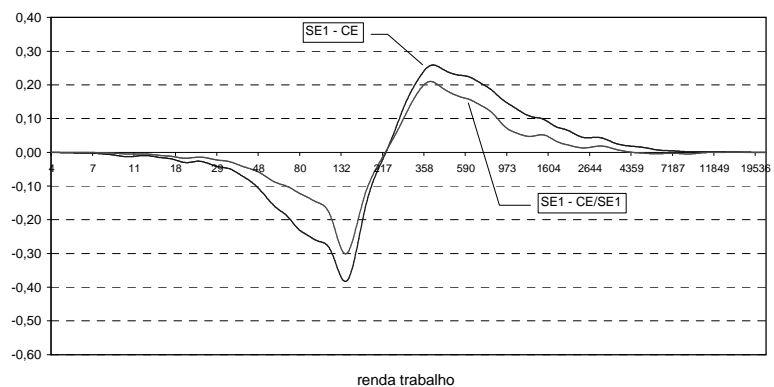
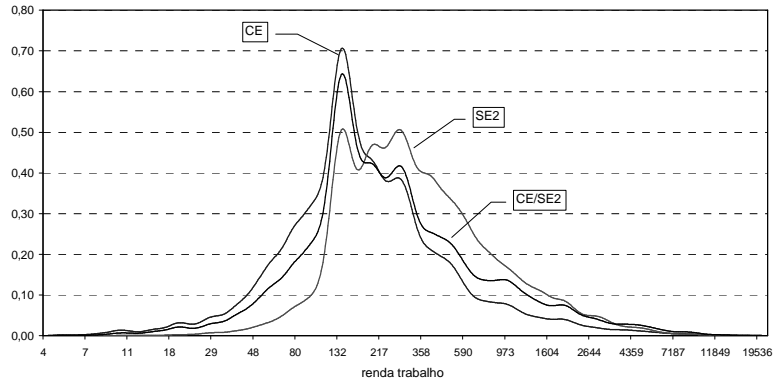


Gráfico A.3

Densidades reais para SE2 e CE e contrafactual para CE com escolaridade de SE2



Diferenças entre as distribuições de SE2 e CE (real e contrafactual)

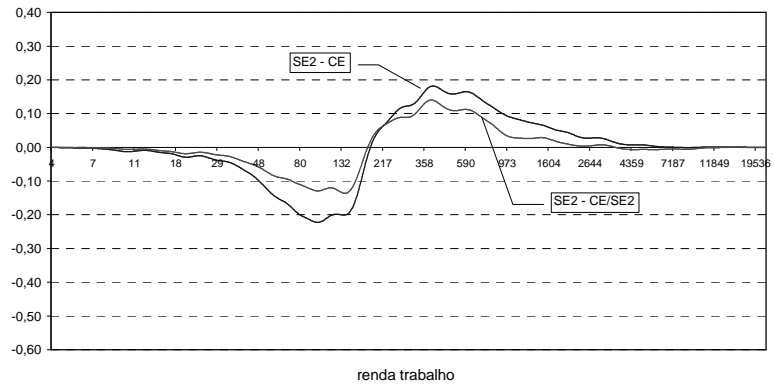


Gráfico A.4 - Evolução da Renda do Trabalho por Percentis para CE, SE1 e CE/SE1

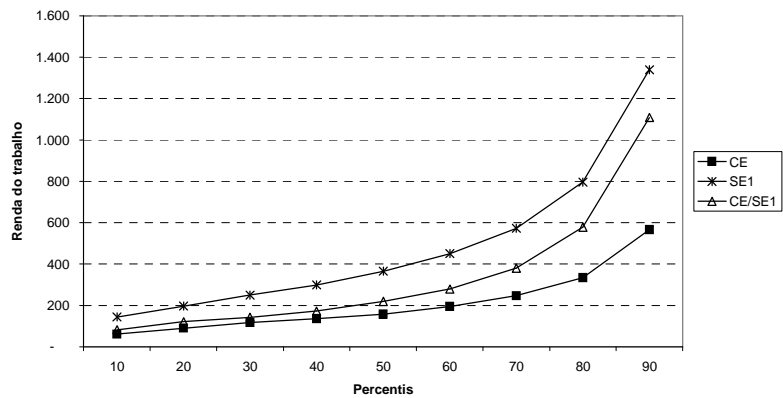


Gráfico A.5 - Evolução da Renda do Trabalho por Percentis para CE, SE2 e CE/SE2

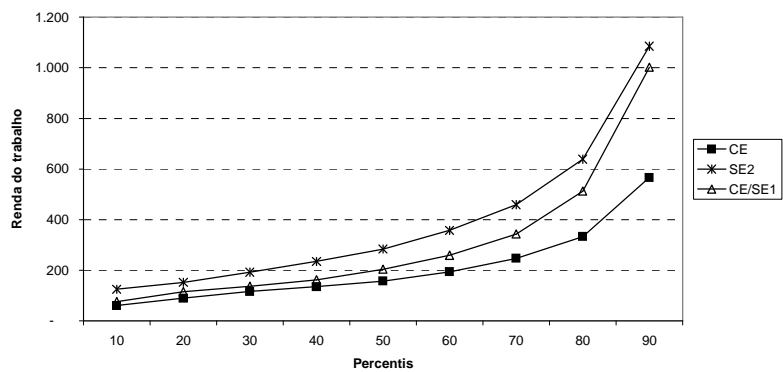


Gráfico A.6 - Evolução da Renda do Trabalho por Percentis para NE, SP e NE/SP

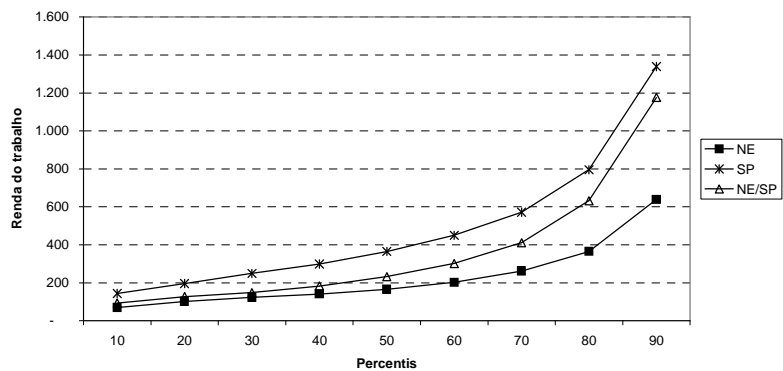


Tabela A.1 - Pontos Notáveis das Distribuições de Renda do Trabalho Estimadas

Percentil	Distribuições Estimadas(*)										
	Brasil	CE	SP	NE	SE1	SE2	CE/SP	CE/SE1	CE/SE2	NE/SP	NE/SE1
10	116	61	187	70	144	125	87	82	77	94	89
20	147	90	250	101	196	153	125	122	116	128	124
30	189	117	302	124	250	192	147	143	137	148	144
40	237	136	358	141	299	235	183	172	162	183	174
50	293	158	428	166	365	284	233	219	204	233	219
60	365	194	523	202	450	358	296	279	260	302	281
70	469	247	672	262	572	459	412	380	344	412	384
80	665	334	916	365	796	639	639	578	513	633	578
90	1.130	567	1.510	639	1.339	1.086	1.200	1.108	1.002	1.176	1.086
Média	572	338	775	358	676	553	584	547	503	569	534
DP	905	697	1.056	704	977	855	1.091	1.041	978	1.033	687
Diferença de Percentis											
10-90	1.014	506	1.323	569	1.195	961	1.113	1.025	926	1.082	997
10-50	177	97	242	96	221	159	145	137	128	139	130
50-90	837	409	1.082	473	974	801	967	888	798	943	867
20-80	518	244	666	264	600	486	514	457	397	505	454

(\*) Os valores apresentados foram extraídos das distribuições estimadas (em log) tomando-se a exponencial de cada ponto.

A média e o desvio-padrão foram calculados diretamente da amostra reponderada com a renda em nível.

Tabela A.2 - Relação da Renda nos Percentis

Percentil	CE x SP	CE/SP x SP	CE x SE1	CE x SE2	CE/SE1 x SE1	CE/SE2 x SE2	NE x SP	NE/SP x SP	NE x SE1	NE x SE2	NE/SE1 x SE1
10	33%	47%	42%	49%	57%	61%	38%	50%	49%	56%	62%
20	36%	50%	46%	59%	62%	76%	41%	51%	52%	66%	63%
30	39%	49%	47%	61%	57%	71%	41%	49%	50%	64%	58%
40	38%	51%	45%	58%	58%	69%	39%	51%	47%	60%	58%
50	37%	54%	43%	55%	60%	72%	39%	54%	45%	58%	60%
60	37%	57%	43%	54%	62%	73%	39%	58%	45%	57%	63%
70	37%	61%	43%	54%	66%	75%	39%	61%	46%	57%	67%
80	36%	70%	42%	52%	73%	80%	40%	69%	46%	57%	73%
90	38%	79%	42%	52%	83%	92%	42%	78%	48%	59%	81%

Tabela A.3 - Coeficientes de Gini e Theil

			Coeficiente de Gini	Coeficiente de Theil
Brasil			0,551	0,607
Regiões	Ponderação Original	CE	0,582	0,768
		NE	0,574	0,731
		SE1	0,523	0,539
		SE2	0,534	0,573
		SP	0,505	0,498
	Reponderação pela Educação	CE/SP	0,622	0,797
		CE/SE1	0,621	0,805
		CE/SE2	0,618	0,810
		NE/SP	0,609	0,760
		NE/SE1	0,608	0,766
	NE/SE2	0,604	0,768	

Obs.: SE1 = MG + RJ + SP

SE2 = MG + RJ

Tabela A.4 - Percentual explicado pela Escolaridade

Métrica	SP x CE	SE1 x CE	SE2 x CE	SP x NE	SE1 x NE	SE2 x NE
Média	56,3%	61,8%	76,7%	50,6%	55,3%	69,2%
Mediana	27,8%	29,7%	36,9%	25,5%	26,9%	32,6%
Coeficiente de Theil	-10,7%	-16,2%	-21,5%	-12,3%	-18,0%	-23,4%
Coeficiente de Gini	-51,9%	-66,8%	-75,0%	-51,4%	-66,9%	-75,0%
Distância de Kullbach-Leibler	55,3%	58,7%	60,3%	51,5%	55,0%	57,3%
Distância de Chernoff	56,0%	59,1%	60,4%	52,3%	55,5%	57,5%
Distância de Sibson	52,5%	56,9%	59,2%	49,2%	53,6%	56,5%