DECOMPONDO O DIFERENCIAL REGIONAL DE SALÁRIOS ENTRE SUDESTE E NORDESTE: UMA APLICAÇÃO DA ABORDAGEM QUANTÍLICA INCONDICIONAL

Vitor Hugo Miro[1]

João Mário Santos de França[2]

Resumo

O presente artigo analisa os diferenciais regionais de salários no Brasil. O método de decomposição com base em regressões RIF, proposto por Firpo *et al*. (2007), foi empregado para decompor as diferenças ao longo das distribuições de salários das regiões Nordeste e Sudeste. As evidências encontradas apontam que as diferenças em termos de características individuais explicam em grande medida o diferencial de rendimentos entre as regiões analisadas, o que corrobora com a hipótese apresentada na literatura mais recente. No entanto, os resultados obtidos com o exercício de decomposição também atribuem grande importância aos fatores institucionais relacionados à formalização e às diferenças na estrutura salarial entre as regiões.

PALAVRAS-CHAVE

Desigualdade regional, diferencias de salários, decomposição, estrutura salarial.

[1] UFC/ Campus de Sobral vitormiro@gmail.com .
[2] CAEN/ UFC. joao.franca@ufc.br .

1. INTRODUÇÃO

O presente artigo investiga o hiato salarial entre regiões brasileiras baseado na ideia de que estas diferenças podem ser explicadas por disparidades na composição de características dos trabalhadores e por diferenças nos retornos (preços) associados a estas características. Tendo como o foco as diferenças salariais entre as regiões Sudeste e Nordeste, a análise aplica um método de regressão quantílica incondicional e decompõe o diferencial em diferentes decis da distribuição dos salários, seguindo o método proposto por Firpo *et al*. (2007).

Na literatura econômica, duas abordagens clássicas se apresentam na explicação das disparidades de renda e rendimentos entre regiões de um mesmo país. Uma delas postula a existência de segmentação dos mercados de trabalho regionais em função da heterogeneidade no desenvolvimento. Outra argumenta que, em equilíbrio, os diferenciais regionais de salários são explicados pela existência de prêmios compensatórios baseados nas diferenças de custo de vida e fatores não-pecuniários, próprios de cada região. Nesse contexto, desvios do equilíbrio podem surgir em decorrência de choques de demanda (Blackaby e Manning, 1990).

No Brasil, predomina na literatura recente abordagens baseadas na teoria de capital humano, em que as desigualdades regionais traduzem simplesmente as desigualdades de renda individuais que, por sua vez, são determinadas pelos atributos produtivos da força de trabalho em cada região, principalmente educação. Essa abordagem pode ser identificada, por exemplo, nos textos de Pessoa (2001), Salvato *et al*. (2010) e Barros (2011). Abordagens alternativas, que apontaram para a importância das heterogeneidades regionais, podem ser vistas em Savedoff (1990) e Menezes e Azzoni (2006).

Na literatura internacional, alguns estudos empíricos adotaram a ideia de que não apenas as distribuições de características individuais seriam importantes para explicar as diferenças regionais, mas também a forma como tais características são remuneradas em cada região. Farber e Newman (1989) sugerem que o estudo apropriado das diferenças regionais de rendimentos deve tratar os salários como o produto de um vetor de características e seus respectivos preços, permitindo a distinção entre fatores permanentes de longo prazo e choques estocásticos de curto prazo. Seguindo essa abordagem, López-Bazo e Montellón (2012) realizaram uma decomposição de diferenciais de rendimentos por regiões na Espanha; Pereira e Galego (2014) estudaram as diferenças regionais de salários entre regiões em Portugal; e Herrera-Idárraga *et al*. (2015) aplica o método de decomposição para avaliar o quanto do diferencial regional de salários na Colômbia podem ser explicados pelas diferenças de escolaridade e formalização no mercado de trabalho.

No presente estudo, a estratégia empírica consiste em avaliar o diferencial de rendimentos do trabalho entre as regiões por um método de decomposição microeconométrica, empregando modelos de equações de rendimentos para a média e para diferentes quantis da distribuição destes rendimentos. Dessa forma os diferenciais de salários são decompostos em termos das diferenças regionais na distribuição de atributos produtivos, como escolaridade, características demográficas e alocação setorial, e em termos de diferenças na estrutura de rendimentos (preços ou retornos destes atributos). Para isso, aplica-se o modelo de decomposição com base em regressões quantílicas incondicionais proposto por Firpo *et al*. (2007 e 2009), também apresentado em Fortin *et al*. (2011), que generaliza a abordagem tradicional de decomposição da média estabelecida por Oaxaca (1973) e Blinder (1973).

Dadas as possibilidades da metodologia adotada, os resultados mostram que as diferenças nos atributos produtivos de trabalhadores nas regiões Nordeste e Sudeste possuem um elevado poder de explicação sobre a desigualdade de salários entre as duas regiões. Esse resultado é bastante condizente com as hipóteses recentes aceca dos diferenciais regionais de renda baseados nas diferenças de capital humano entre as regiões. Porém, a contribuição das diferenças na *estrutura salarial* também se mostrou relevante, com destaque para os retornos educacionais, o que incorpora novas evidências ao estudo das desigualdades regionais.

Além dessa introdução o artigo está organizado da seguinte forma. A seção dois apresenta uma breve revisão da literatura. Na terceira seção tem-se uma análise descritiva com base na amostra empregada no trabalho, destacando os diferenciais na distribuição de salários entre o Sudeste e o Nordeste, e ressaltando as principais diferenças de características dos trabalhadores de ambas regiões. Grande ênfase é dada à quarta seção, que apresenta a metodologia de análise adotada com base na decomposição e no método de regressões quantílicas incondicionais. A quinta seção apresenta e discute os resultados da decomposição, enquanto a sexta e última seção expõe as considerações finais do trabalho.

2. REVISÃO DA LITERATURA

De modo geral, a literatura que se dedicou a estudar e discutir diferenças regionais de salários são guiadas por algumas ideias principais. A mais clássica delas, afirma que em mercados de trabalho relativamente homogêneos e com mobilidade do fator trabalho, as diferenças salariais entre regiões podem existir como forma de equilibrar diferenças em termos do custo de vida e em atributos não-pecuniários que afetam a utilidade de trabalhadores e lucros das firmas (Blackaby e Manning, 1990). O modelo competitivo que ampara essas suposições sugere que o retorno de características dos trabalhadores tende a convergir entre diferentes

regiões na ausência de informações imperfeitas, barreiras à mobilidade e distúrbios estocásticos persistentes (Farber and Newman, 1989).

Abordagens alternativas, mesmo que derivadas da abordagem competitiva clássica, também possuem um papel importante na literatura. Farber e Newman (1989) recorre ao modelo de *salário eficiência* para justificar diferenças salariais entre trabalhadores idênticos. Considerando que esse modelo oferece uma explicação para a persistência de diferenciais interindustriais de salários e, assumindo que a estrutura industrial difere entre regiões, potencialmente tem-se diferenças regionais de salários.

Sob o ponto de vista empírico, em um modelo simples de determinação de salários existem dois determinantes primários para as diferenças regionais de salários: i) diferenças na distribuição de características individuais, de trabalhadores e firmas, e ii) a forma como estas características são remuneradas em cada região, ou seja, a estrutura salarial.

Estudando as diferenças salariais entre regiões americanas, Gerking e Weirick (1983) concluem que os diferenciais são explicados exclusivamente por diferenças nas características. Por outro lado, Farber e Newman (1987) sugerem que as diferenças nos retornos podem ser tão importantes quanto a heterogeneidade da força de trabalho. Conclusão similar para o diferencial salarial entre Norte e Sul no Reino Unido foi obtida por Blackaby e Manning (1990).

Para compreender o papel desses dois determinantes, muitos estudos adotaram, além da estimação de equações de rendimentos, métodos de decomposição. Motellón *et al*. (2011) analisa o diferencial regional na Espanha adotando a metodologia proposta por Dinardo *et al*. (1996). López-Bazo e Montellón (2012) abordam as diferenças de salários entre regiões da Espanha empregando o método de decomposição com correção para viés de seleção sugerido por Neuman e Oaxaca (2004). Pereira e Galego (2014) estudaram as diferenças regionais de salários entre regiões em Portugal empregando o modelo de decomposição baseado em regressões quantílicas sugerido inicialmente por Machado e Mata (2005). Herrera-Idárraga *et al*. (2015) aplica o método de decomposição por quantis incondicionais proposto por Firpo *et al*. (2007) para avaliar o quanto do diferencial regional de salários na Colômbia podem ser explicados pelas diferenças de escolaridade e formalização no mercado de trabalho.

No Brasil, predomina na literatura recente abordagens baseadas na teoria de capital humano, em que as desigualdades regionais são determinadas basicamente pela desigualdade na distribuição de atributos produtivos da força de trabalho em cada região.

Savedoff (1990) analisa dados das décadas de 1970 e 1980 e argumenta que explicações tradicionais para os diferenciais regionais de salário no Brasil, como as diferenças nos custos de vida e na composição da força de trabalho (capital humano), seriam responsáveis por apenas

parte dos diferenciais, uma vez que as disparidades de rendimentos persistem, mesmo controlando-se por características pessoais e ocupacionais e ponderando-se pelas diferenças regionais de custo de vida. Esse autor também descarta que diferenciais compensatórios e barreiras à mobilidade de mão-de-obra possam explicar grande parcela do diferencial. Para o autor, os fatores que potencialmente explicam a persistência dos diferenciais regionais de rendimentos estão relacionados ao dinamismo da demanda por trabalho em cada região, o que inclui segmentações intra-regionais derivadas da organização setorial e estratificação do mercado de trabalho.

O texto de Pessoa (2001) propõe uma reflexão sobre o problema da desigualdade regional de renda no Brasil. Sob a consideração das disparidades de renda *per capita*, o autor argumenta que na análise das desigualdades regionais, deve-se considerar a distinção entre dois problemas de natureza distinta. No primeiro a baixa renda *per capita* de uma região em relação às demais é consequência de características embutidas na região. No segundo caso, a renda de determinada região é baixa por características dos moradores daquela região. A distinção empírica entre esses problemas poderia ser feita estudando os diferenciais de renda entre trabalhadores de diferentes regiões, controlando pelas características desses trabalhadores. Para um dado diferencial de renda *per capita* entre as regiões, o problema regional será predominante se os diferenciais de renda entre os trabalhadores com as mesmas características forem persistentes. De outra forma, se após o controle das características dos trabalhadores o diferencial regional de renda reduzir-se significativamente, pode-se inferir que o problema é correlacionado com a região, mas não se trata de um problema genuinamente regional, e sim social.

Adotando uma hipótese alternativa, Menezes e Azzoni (2006) buscam identificar em quanto características regionais contribuem na explicação do diferencial regional de salários. Analisando a convergência de salários entre regiões metropolitanas, eles apontam que diferenças no aparato produtivo entre as regiões, traduzidas principalmente na demanda por trabalho, também são importantes para explicar as disparidades de salários e sua persistência ao longo do tempo.

O estudo de Salvato *et al*. (2010) buscou identificar o quanto o diferencial de renda do trabalho entre as regiões Nordeste e Sudeste (e entre os estados do Ceará e São Paulo) é explicado pelo diferencial de escolaridade da população. A análise empírica emprega o método de DiNardo *et al*. (1996) para construir funções de densidade contrafactuais reponderando os indivíduos da região base (Nordeste) pela distribuição de educação da região a ser comparada (Sudeste). Os resultados mostram que a reponderação pela escolaridade eleva a renda média da

região Nordeste em cerca de 50% e que a maior contribuição ocorre no segmento superior da distribuição. Outro fato destacado pelos autores foi que ao estabelecer a distribuição de escolaridade do Sudeste ao Nordeste, mantendo-se a estrutura salarial, a desigualdade de rendimentos torna-se maior.

Barros (2011) adota hipótese semelhante à de Pessoa (2001), afirmando que o que poderia se configurar como um problema regional no Brasil é o relativo atraso observado nos indicadores econômicos e sociais da região Nordeste. O autor defende que a existência de um problema regional se dá quando as oportunidades para indivíduos de características semelhantes não são as mesmas nas diferentes regiões.

3. DADOS E EVIDÊNCIAS PRELIMINARES

Os dados empregados nesse estudo são provenientes da base de microdados da Pesquisa Nacional por Amostras de Domicílios (PNAD), levada a campo anualmente pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Microdados de pesquisas amostrais como a PNAD permitem um controle apropriado das diferenças de distribuição de características individuais entre as regiões.

Foram considerados dados em dois períodos diferentes para a composição de uma subamostra relacionada aos objetivos do estudo aqui proposto, os biênios 2002/2003 e 2012/2013. O objetivo, além de avaliar o diferencial de rendimentos em cada período, é poder observar se nesse intervalo de 10 anos houve alguma mudança significativa que explicasse as diferenças regionais. Essa subamostra incluiu indivíduos economicamente ativos residentes nas duas regiões consideradas, com idade entre 24 e 65 anos, de ambos os sexos, residentes em áreas urbanas e rurais. Entre os ocupados foram considerados indivíduos envolvidos em atividades agrícolas, industriais e dos setores de serviços e comércio. Os ocupados também foram qualificados em empregados formais (com carteira de trabalho assinada), informais (sem carteira de trabalho assinada) ou trabalhadores por conta própria[3]. A Tabela 1 apresenta estatísticas descritivas dessa subamostra em cada um dos períodos para ambas as regiões. Uma breve análise desses dados permite ver diferenças importantes entre a força de trabalho das duas regiões.

[3] Assim como outros estudos na mesma literatura, não foram considerados funcionários públicos e empregadores, em função das características da formação de salários dessas categorias.

Tabela 1: Composição da força de trabalho (em %) – Nordeste e Sudeste (2002/2003 e 2012/2013).

	Nordeste		Sudeste	
	2002/2003	*2012/2013*	*2002/2003*	*2012/2013*
Nº de observações (amostra/PEA)	61.708	58.655	65.414	67.252
Ocupação (% da PEA)	91,0	91,8	89,3	93,8
Escolaridade				
Anos de estudo (média)	4,9	7,0	7,2	8,8
Analf./ Fundam. (1º ciclo) incompleto	43,7	25,7	18,4	10,3
Fundamental (1º ciclo) completo	25,0	22,7	32,6	22,0
Fundamental (2º ciclo) completo	11,1	14,9	17,2	17,8
Médio completo	17,5	31,0	24,4	39,0
Superior completo	2,7	5,7	7,3	11,0
Características demográficas				
Grupo etário (24-29 anos)	24,1	21,6	22,7	21,9
Grupo etário (30-39 anos)	34,1	33,4	33,4	33,4
Grupo etário (40-49 anos)	23,3	25,4	26,7	25,1
Grupo etário (50-65 anos)	18,5	19,6	17,2	19,6
Gênero (% feminino)	37,2	40,3	42,8	43,6
Cor/raça (% pardos| negros| índios)	72,5	74,3	38,7	47,9
Distribuição espacial				
Zona urbana	73,6	78,6	92,5	93,6
Área metropolitana	23,6	25,3	45,9	44,8
Distribuição setorial*				
Agrícola	29,8	15,9	8,5	6,8
Indústria	20,5	21,3	32,0	32,2
Serviços e comércio	49,7	55,1	59,5	60,9
Formal	24,7	41,5	48,1	57,8
Informal	30,6	25,6	24,4	18,3
Conta própria	44,7	29,5	27,4	23,8

Fonte: Estimações dos autores. Microdados do IBGE/PNAD.
*Apenas para ocupados.

Uma vez que a literatura aponta as disparidades de capital humano como o principal determinante para os diferenciais de salários, caracterizar as diferenças entre as regiões é muito relevante para a presente análise. Considerando a amostra analisada, em termos de anos de estudo, a escolaridade média dos indivíduos no Nordeste aumentou de 5 para 7 anos de estudo entre os dois períodos; enquanto que no Sudeste, a média passou de 7 para quase 9 anos.

Na região Nordeste, em 2012/2013, quase 50% da PEA não possuía o ensino fundamental completo, sendo que quase 26% se enquadrava no grupo com escolaridade mais baixa, sem ter concluído pelo menos 4 anos de estudo (o que consideramos como equivalente ao primeiro ciclo do ensino fundamental). No caso da região Sudeste, no mesmo período, o percentual de trabalhadores que não possuíam o ensino fundamental completo era de aproximadamente 32%. Com relação aos níveis de ensino mais elevados, 31% dos nordestinos

haviam concluído o ensino médio e apenas 5,7% concluíram o ensino superior. Na região Sudeste, 39% tinham o ensino médio completo e 11% possuía um diploma universitário.

A Figura 1 a seguir mostra a distribuição acumulada de anos de estudo para as duas regiões. No gráfico apresentado é possível ver as diferenças regionais na proporção acumulada, principalmente no segmento de escolaridade mais baixa, e a forma como essa proporção diminui entre os dois períodos. Um destaque pode ser feito para a semelhança das distribuições acumuladas no Nordeste de 2012/2013 com o Sudeste de 2002/2003.

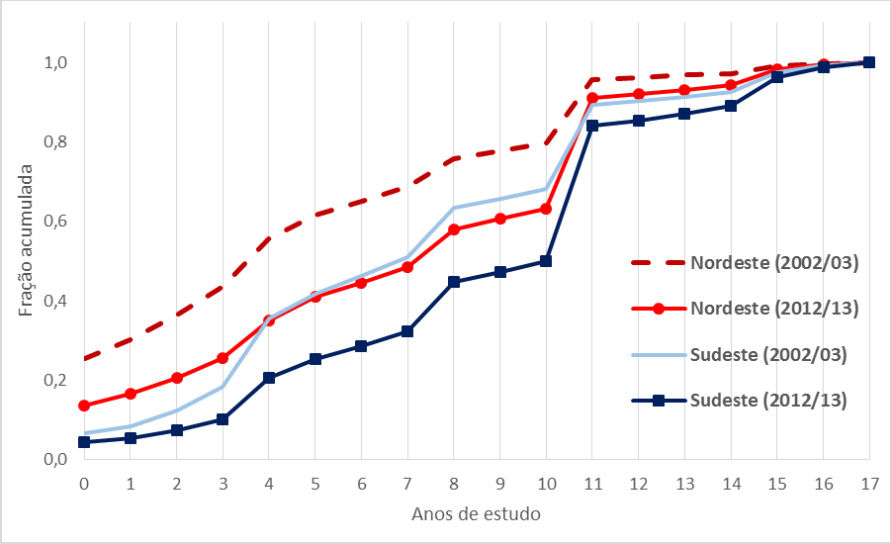


Figura 1: Distribuição acumulada – anos de estudo.
Fonte: Estimativa dos autores. Microdados da PNAD/IBGE.

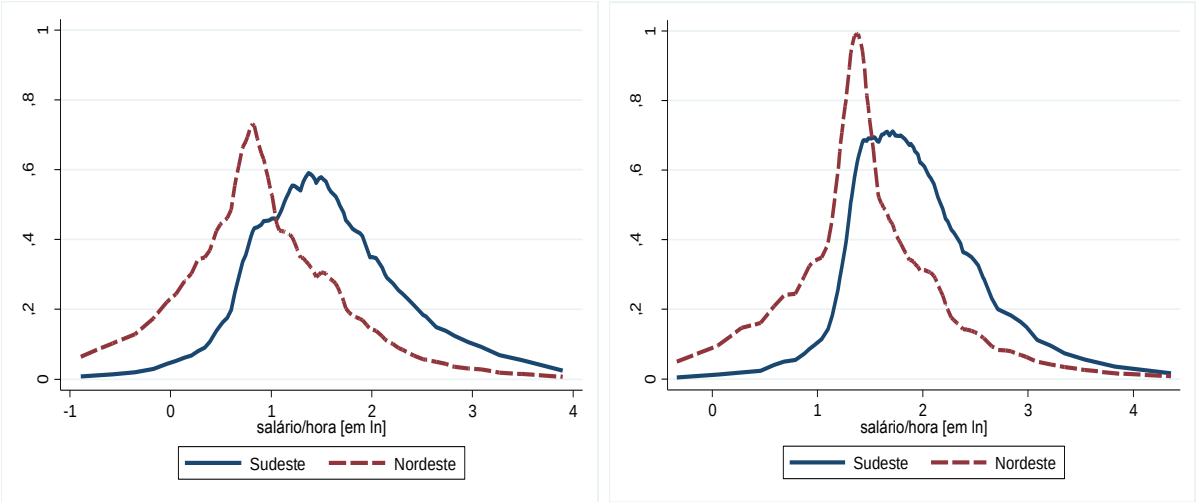
Dentre as características demográficas, não existem diferenças significativas na distribuição etária, e o que pode ser destacado é o relativo envelhecimento entre os dois períodos considerados. Quanto a participação feminina, esta cresceu ao longo dos dez anos do intervalo, mas permanecendo um pouco maior na região Sudeste. A maior diferença está no percentual de indivíduos que se declararam pardos, negros ou de origem indígena, que corresponde a mais de 70% no Nordeste, enquanto que no Sudeste, apesar do aumento, essa proporção foi inferior a 50%.

Diferenças também podem ser observadas na distribuição da força de trabalho entre locais de residência. No Nordeste, em 2012/2013, mais 20% ainda se encontram em áreas rurais enquanto que no Sudeste essa proporção não chega a 7%. Além de ser mais urbana, a população da região Sudeste também é bastante concentrada nas regiões metropolitanas, que concentrava quase 45%, contra aproximadamente 25% no Nordeste (em 2012/2013).

Com relação a distribuição setorial, o setor de serviços é predominante em ambas as regiões, e apresentou crescimento no período considerado (em 2012/2013, 55% no Nordeste e

61% no Sudeste). As maiores diferenças estão no grande peso relativo do setor agrícola na região Nordeste (16% contra 6,8% no Sudeste em 2012/2013) e o do setor industrial no Sudeste (32% contra 21% no Nordeste, em 2012/2013). Com respeito a formalização, observa-se uma tendência de aumento da formalização, com destaque para o crescimento no Nordeste. Mas ainda assim, em 2012/2013, o Sudeste apresentava uma taxa de formalização bem superior à da região Nordeste (57,8% frente 41,8% em 2012/2013).

Com o objetivo de estudar o diferencial de salários, emprega-se a medida do logaritmo (natural) do salário/hora, em que foram considerados os valores referentes aos rendimentos do trabalho único ou principal dos indivíduos, em valores reais[4]. O gráfico da Figura 2 mostra funções de densidade estimadas dessa variável de rendimentos para as duas regiões e nos dois biênios considerados.



2002/2003 **2012/2013**

Figura 2: Densidades estimadas (*ln* do salário/hora).
Fonte: Estimativa dos autores. Microdados da PNAD/IBGE.
*Função de Epanechnikov com intervalo de classe de 0,065.

Os gráficos da Figura 3 destacam o comportamento dos salários e os diferenciais segundo decis da distribuição. No gráfico superior, é possível ter uma ideia da evolução dos salários ao longo dos decis e a forma como eles evoluíram na década considerada. Uma visualização mais clara dos diferenciais regionais de salários é possível no gráfico inferior, em que é possível notar uma diferença maior nos decis inferiores e certa estabilidade entre os anos considerados.

[4] Os valores foram atualizados para setembro/outubro de 2013 de acordo com o Índice Nacional de Preços ao Consumidor (INPC) também divulgado pelo IBGE.

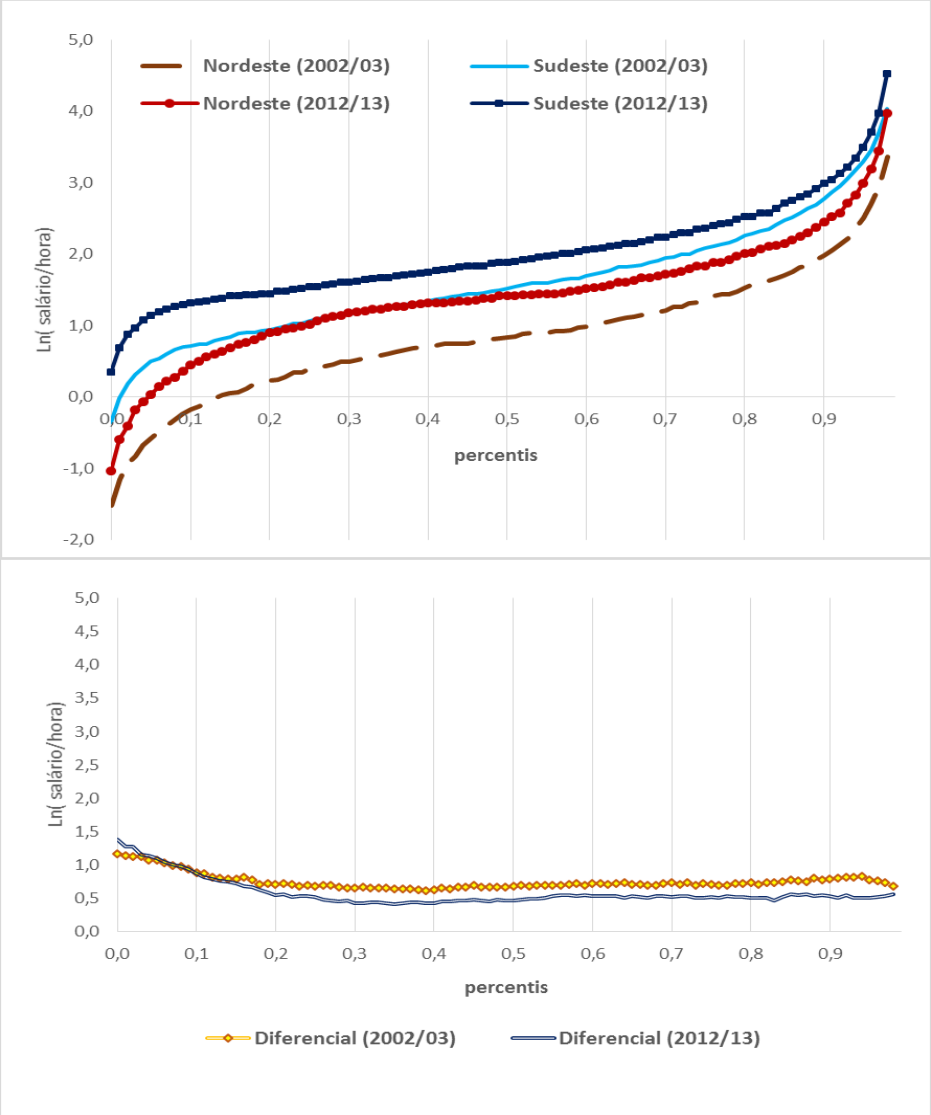


Figura 3: Diferencial de rendimentos do trabalho (*ln* do salário/hora).
Fonte: Estimativa dos autores. Microdados da PNAD/IBGE.

4. METODOLOGIA

Nesta seção descreve-se o método empírico empregado na estimação das equações de rendimentos e na decomposição dos diferenciais de salários entre as regiões. Tendo como base o modelo de equações de rendimentos mincerianas[5], a forma mais simples e popular de obter uma medida do diferencial de rendimentos entre as regiões é estimar uma única equação incorporando *dummies* para cada região. Essa abordagem, no entanto, impõe que os retornos das demais características individuais sejam idênticos entre as regiões. Essa suposição pode ser bastante restritiva, uma vez que exista segmentação regional, trabalhadores com as mesmas características obterão retornos diferenciados entre as regiões.

[5] MINCER, J..***Schooling, Experience and Earnings***. National Bureau of Economic Research, 1974.

Uma forma alternativa, e mais adequada de lidar com essa questão, seria estimar uma equação de rendimentos para cada região, de forma que o salário de um indivíduo i na região r seja dado por:

$$y_{ir} = X_{ir}\beta_r + \varepsilon_{ir} \qquad [1]$$

Onde y_{ir} denota o logaritmo do salário/hora do indivíduo i na região r; o termo X_{ir} é composto por um conjunto de características que supostamente afetam o salário desse indivíduo, incluindo educação, experiência, gênero, cor/raça, setor, entre outras. O coeficiente β_r é o vetor de retornos associados às características X_{ir}. E ε_{ir} é um termo de erro idiossincrático.

Na aplicação do método clássico de Oaxaca-Blinder (Oaxaca, 1973 e Blinder, 1973) para a decomposição da média, a equação [1] é estimada para cada região, o procedimento de análise do diferencial de rendimentos é realizado com a aplicação da decomposição. No presente estudo utiliza-se uma generalização do método de Oaxaca-Blinder, proposta por Firpo *et al*. (2007). Esse método emprega regressões com base na função de influência recentrada (*RIF*) combinadas com um procedimento de reponderação inspirado em DiNardo *et al*. (1996).

4.1. O método de regressão por quantis incondicionais – o método RIF

Em geral qualquer estatística da distribuição de uma variável de interesse y, uma medida de posição, dispersão ou desigualdade, pode ser escrita como um funcional $v(F_y)$ da função de distribuição $F_y(y)$. Firpo *et al*. (2006; 2009) apresenta uma abordagem que permite a estimação do efeito de mudanças em um conjunto de covariadas X sobre a estatística $v(F_y)$ da distribuição marginal de y. O método se baseia em regressões recentradas na função de influência – denominada como *RIF* (do inglês *recentered influence function*).

A função de influência de uma estatística é essencialmente a sua derivada (direcional) de primeira ordem. Denotada por $IF(y; v, F_y)$, a função de influência de $v(F_y)$ mede o efeito relativo (a influência) de uma pequena perturbação na distribuição de y sobre o resultado da estatística de interesse. Considerando a média, $\mu(F_y)$, a função de influência é dada por $IF(y; \mu, F_y) = y - \mu(F_y)$. Para o caso do τ-ésimo quantil da distribuição, definido como $q_\tau(F_y) = inf\{y|F(y) \geq \tau\}$, a função de influência é dada por: $IF(y; q_\tau, F_y) = \frac{\tau - \mathbb{I}\{y \leq q_\tau(F_y)\}}{f_y(q_\tau(F_y))}$; onde $\mathbb{I}\{y \leq q_\tau(F_y)\}$ é um função indicadora se a variável resposta é menor ou igual ao quantil q_τ, e $f_y\left(q_\tau(F_y)\right)$ é a função de densidade marginal de y avaliada em q_τ.

A função de influência recentrada é definida como a soma da estatística da distribuição e a sua função de influência, ou seja: $RIF(y; \upsilon, F_y) = \upsilon(F_y) + IF(y; \upsilon, F_y)$.

Para a média $\mu(F_y)$, a *RIF* é dada simplesmente por y, enquanto que para o τ-ésimo quantil, a RIF é dada por:

$$RIF(y; q_\tau, F_y) = q_\tau + \frac{\tau - \mathbb{I}\{y \leq q_\tau(F_y)\}}{f_y\left(q_\tau(F_y)\right)}$$

$$= c_{1\tau} \cdot \mathbb{I}\{y > q_\tau\} + c_{2\tau} \qquad [\,2\,]$$

onde $c_{1\tau} = \frac{1}{f_y(q_\tau)}$ e $c_{2\tau} = q_\tau - c_{1\tau} \cdot (1 - \tau)$.

Uma vez que o valor esperado da função de influência de qualquer estatística da distribuição é, por definição, igual a zero; o valor esperado da *RIF* correspondente é igual a própria estatística, ou seja, $\mathbb{E}[RIF(y; \upsilon, F_y)] = \upsilon(F_y)$. [6]

Assumindo que y é observado em associação a um conjunto de covariáveis (X), pode-se modelar a expectativa condicional da $RIF(y; \upsilon, F_y)$ em função de X $\mathbb{E}[RIF(y; \upsilon, F_y)|X = x]$. Essa expectativa condicional é denominada como *regressão RIF* e pode ser analisada como uma regressão quantílica incondicional (Firpo *et al*., 2006 e 2009).

Assumindo que a *RIF* pode ser aproximada por uma função linear em termos de X, tem-se $RIF(y; \upsilon, F_y|X) = X\beta + \varepsilon$. Aplicando a Lei de Expectativas Iteradas e considerando a hipótese de que $\mathbb{E}[\varepsilon|X] = 0$, tem-se:

$$\upsilon(F_y) = \mathbb{E}_X\left[\mathbb{E}[RIF(y; \upsilon, F_y)]\right] = \mathbb{E}[X] \cdot \beta \qquad [\,3\,]$$

E dessa forma os parâmetros em β podem ser estimados por método de regressão. Do ponto de vista prático, a estimação da regressão *RIF* para a média da distribuição pode ser implementada de forma simples e direta por métodos de regressão linear. No entanto, no caso de quantis da distribuição a implementação não é tão direta. O primeiro passo consiste em computar os quantis amostrais $\hat{q}_\tau$ e estimar a função de densidade $\hat{f}_y(\hat{q}_\tau)$, para então computar a *dummy* $\mathbb{I}\{y \leq q_\tau\}$. Com isso, a estimativa para cada observação da $\widehat{RIF}(y, q_y)$ é obtida

[6] Uma propriedade importante da função de influência é que, sempre que a frequência e a amplitude dos valores de y são limitados, seu valor esperado é nulo, ou seja, $\int IF(y; \upsilon)dF(y) = 0$. Dessa forma $\int RIF(y; \upsilon)dF(y) = \int [\upsilon(F) + IF(y, \upsilon)] \cdot dF(y) = \upsilon(F)$.

empregando a equação [2]. O segundo passo é estimar a regressão $RIF(y, q_y)$ sobre o vetor X, o que pode ser feito aplicando o método de MQO[7].

Os parâmetros estimados pela regressão *RIF* possuem uma interpretação condicional e não condicional. No caso da média, a estimação da regressão *RIF* é equivalente a estimação por MQO. Já a estimação da regressão *RIF* para quantis não é equivalente à estimação quantílica condicional, como a proposta por Koenker e Basset (1978).

No contexto de uma equação de rendimentos, a regressão quantílica condicional provê estimativas do retorno de características individuais, *coeteris paribus*, de forma que esse retorno é variável entre os indivíduos de acordo com o quantil condicional ao qual ele pertence. Por outro lado, regressões quantílicas incondicionais estimam, *coeteris paribus*, o efeito de pequenas mudanças em uma característica dos indivíduos em cada quantil da distribuição, o que permite avaliar o efeito sobre uma variada gama de estatísticas da distribuição dos rendimentos (Fournier e Koske, 2012).

4.2. O método de decomposição

Com foco na decomposição de diferenças na distribuição de rendimentos entre duas regiões, definidas genericamente como A e B, aplica-se um método de decomposição proposto por Firpo *et al*. (2007). Generalizando a decomposição de Oaxaca-Blinder, pode-se representar a diferença nas distribuições de rendimentos de duas regiões por meio de alguma estatística dessas distribuições. Assim, o diferencial total entre as distribuições é escrito como:

$$\Delta^{\upsilon} = \upsilon(F_{y_B}) - \upsilon(F_{y_B}) \qquad [4]$$

onde $\upsilon(F_{y_r})$, da mesma forma como definido anteriormente, representa uma estatística da distribuição de rendimentos para a região $r = A, B$.

Um método de decomposição típico irá dividir Δ^{υ} em um componente associado a diferenças de características observáveis dos indivíduos e um componente associado a diferenças na *estrutura salarial* entre as duas regiões. A "chave" para esse procedimento é a construção de uma distribuição contrafactual (F_{y_C}) e, obviamente, da respectiva estatística de interesse $\upsilon(F_{y_C})$. Pode-se então simular uma distribuição de rendimentos compondo a *estrutura*

[7] Firpo *et al*. (2009) apresentam métodos alternativos para implementar a estimação. No entanto, em termos práticos as estimativas são muito próximas, não se observando uma grande vantagem no uso de métodos mais sofisticados.

salarial da região A e a distribuição de características (observáveis e não observáveis) dos trabalhadores da região B.

Empregando esse contrafactual e algumas manipulações algébricas pode-se obter:

$$\Delta^{\upsilon} = \left[\upsilon\left(F_{y_B}\right) - \upsilon\left(F_{y_C}\right)\right] + \left[\upsilon\left(F_{y_C}\right) - \upsilon\left(F_{y_A}\right)\right]$$

$$\Delta^{\upsilon} = \Delta_S^{\upsilon} + \Delta_X^{\upsilon} \qquad [\ 5\ ]$$

O primeiro termo, Δ_S^{υ}, reflete o efeito das diferenças na *estrutura salarial*, determinada pelos retornos (ou preços) das características consideradas; comumente denominada como *efeito preço* ou da *estrutura salarial*. O termo Δ_X^{υ} representa as diferenças nas distribuições de características, normalmente denotado como *efeito quantidade* ou *efeito composição*.

Conforme apresentado anteriormente, tomando o valor esperado das funções de influência recentradas e assumindo uma especificação linear,

$$\upsilon\left(F_{y_r}\right) = \mathbb{E}[RIF(y_r; \upsilon_r)|X, R = r] = X_r\beta_r \qquad \text{para } r = A, B \qquad [\ 6\ ]$$

para as distribuições observadas e,

$$\upsilon\left(F_{y_C}\right) = \mathbb{E}[RIF(y_A; \upsilon_C)|X, R = B] = X_C\beta_C \qquad [\ 7\ ]$$

para a distribuição contrafactual.

Aplicando a decomposição de OB sob a suposição de linearidade, a expectativa da estatística contrafactual em que se adota a distribuição de características da região B e a *estrutura salarial* da região A é estimada de forma que $\bar{X}_C\hat{\beta}_C = \bar{X}_B\hat{\beta}_A$. No entanto, quando a linearidade não se mantém, essa igualdade não é válida. No contexto da decomposição de Oaxaca para a média, Barsky *et al*. (2002) demonstra que quando a verdadeira esperança condicional não é linear, a decomposição com base em uma regressão linear pode ser tendenciosa. Os autores argumentam que uma das possíveis soluções para esse problema é adotar uma abordagem de reponderação como a proposta por DiNardo *et al*. (1996).

Considerando a sugestão de Barsky *et al*. (2002), Firpo *et al*. (2007) propõe então uma abordagem que combina o método de regressões RIF com reponderações inspiradas na técnica apresenta por DiNardo *et al*. (1996). Essa abordagem estabelece o uso de um fator de reponderação, que pode ser definido por:

$$\psi(X) = \frac{Pr[R = B|X]}{Pr[R = A|X]} \cdot \frac{Pr[R = A]}{Pr[R = B]} \qquad [\ 8\ ]$$

onde $Pr[R = r|X]$ é a probabilidade de um trabalhador pertencer a região r dados as suas características X e $Pr[R = r]$ denota a proporção de trabalhadores na região r. Em termos práticos, esse fator de ponderação pode ser computado pela estimação de um modelo de probabilidade para $Pr[R = B|X]$, e usando as probabilidades preditas para calcular o valor de $\hat{\psi}(X)$ para cada observação.

Estimando regressões RIF para as regiões A, B por MQO, tem-se:

$$\hat{\beta}_r = \left(\sum_{i \in r} \widehat{\omega}_r \cdot X_i \cdot X_i'\right)^{-1} \cdot \sum_{i \in r} \widehat{\omega}_r \cdot \widehat{RIF}(y_{ri}; \upsilon_r) \cdot X_i \quad , para\ r = A, B \qquad [\ 9\ ]$$

onde $\widehat{\omega}_r$ é um fator de ponderação baseado na própria composição da amostra total. E para o contrafactual, considerando o fator de reponderação, estima-se:

$$\hat{\beta}_C = \left(\sum_{i \in A} \hat{\psi}(X_i) \cdot X_i \cdot X_i'\right)^{-1} \cdot \sum_{i \in A} \hat{\psi}(X_i) \cdot \widehat{RIF}(y_{Ai}; \upsilon_C) \cdot X_i \qquad [\ 10\ ]$$

$$\bar{X}_C = \sum_{i \in A} \hat{\psi}(X_i) \cdot X_i$$

onde $plim(\bar{X}_C) = plim(\bar{X}_B)$. Se a expectativa condicional da $RIF(y_A; \upsilon_C)$ dado X é linear, ambas as regressões ponderada e reponderada irão render a mesma estimativa de β_A, ou seja, $plim(\hat{\beta}_C) = plim(\hat{\beta}_A) = \beta_A$. Quando essa expectativa condicional não é linear, as estimativas geralmente irão divergir.

A decomposição em termos dessas estimativas é dada por:

$$\widehat{\Delta}^{\upsilon} = \left[\bar{X}_B \hat{\beta}_B - \bar{X}_C \hat{\beta}_C\right] + \left[\bar{X}_C \hat{\beta}_C - \bar{X}_A \hat{\beta}_A\right]$$

$$\widehat{\Delta}^{\upsilon} \ = \ \widehat{\Delta}_S^{\upsilon} \ + \ \widehat{\Delta}_X^{\upsilon} \qquad [\ 11\ ]$$

A estimativa do *efeito composição* $\widehat{\Delta}_X^{\upsilon}$ pode ser dividida em dois componentes em que um termo padrão $(\bar{X}_C - \bar{X}_A)\hat{\beta}_A$ é somado a um erro de especificação $\bar{X}_C(\hat{\beta}_C - \hat{\beta}_A)$, de forma que :

$$\widehat{\Delta}_X^{\upsilon} \ = (\bar{X}_C - \bar{X}_A)\hat{\beta}_A + \bar{X}_C(\hat{\beta}_C - \hat{\beta}_A) \qquad [\ 12\ ]$$

Se a especificação linear do modelo é correta, o erro de especificação tende a ser nulo; uma vez que $plim(\hat{\beta}_C) = plim(\hat{\beta}_A)$.

Em termos de uma decomposição detalhada, em que é possível estimar a contribuição de cada variável explicativa $k = 1, \cdots, K$, o *efeito composição* pode ser escrito como:

$$\widehat{\Delta}_X^{\upsilon} = \sum_{k=1}^{K} (\bar{X}_{Ck} - \bar{X}_{Ak})\hat{\beta}_A \qquad [\ 13\ ]$$

O efeito da *estrutura salarial* $\widehat{\Delta}_S^{\upsilon}$, por sua vez, passa a ser definido por:

$$\hat{\Delta}_S = \bar{X}_B(\hat{\beta}_B - \hat{\beta}_C) + (\bar{X}_B - \bar{X}_C)\hat{\beta}_C \qquad [\ 14\ ]$$

O efeito da *estrutura salarial* se reduz ao primeiro termo, uma vez que, considerando grandes amostras e a estimação correta do fator de reponderação, o erro de reponderação tende a ser nulo, uma vez que, se o fator de reponderação é estimado de forma consistente, ou seja, $plim(\bar{X}_C) = plim(\bar{X}_B) \Rightarrow plim(\bar{X}_B - \bar{X}_C) = 0$.

O efeito da *estrutura salarial* também pode ser reescrito em termos de uma decomposição detalhada,

$$\widehat{\Delta}_X^{\upsilon} = (\hat{\beta}_{B1} - \hat{\beta}_{C1}) + \sum_{k=2}^{K} \bar{X}_{Bk}(\hat{\beta}_{Bk} - \hat{\beta}_{Ck}) \qquad [\ 15\ ]$$

Nessa expressão o primeiro elemento corresponde à diferença entre os interceptos das equações para A e B, compondo uma espécie de efeito residual. O segundo termo representa a contribuição dos retornos de cada variável explicativa. Tanto a diferença residual, quanto o efeito da *estrutura salarial* para cada covariada dependem diretamente da escolha do grupo omitido no processo de estimação. Trata-se de uma das deficiências do método de decomposição de OB e que, como argumentado por Fortim *et al*. (2011), não apresenta uma solução geral satisfatória[8].

Uma ressalva importante sobre o método de regressões quantílicas incondicionais fornecem uma estimativa de equilíbrio parcial da variável de interesse, assumindo que mudanças marginais em X não exercem impacto sobre a distribuição conjunta de X e y, significando que as taxas de retorno não variam em resposta a qualquer variação pequena em uma das características em X. O método também assume a forte hipótese de independência entre a heterogeneidade não observada e as características observadas. Embora estas suposições não se sustentem na prática, Fournier e Kosque (2012) argumentam que uma comparação entre

[8] O problema do "grupo omitido" é comum no caso de variáveis explicativas categóricas. Possíveis correções para essa questão são apontadas por Oaxaca e Ransom (1999) e Yun (2005, 2008).

estimativas de diferentes quantis permanece válida, desde que o viés potencial seja o mesmo em toda a população da amostra.

5. RESULTADOS

5.1. Regressões RIF

Foram estimadas regressões RIF para o salário médio e decis da distribuição de rendimentos para as regiões Sudeste e Nordeste nos dois biênios considerados[9]. Os coeficientes estimados para estas regressões (de acordo com a expressão [9]), são apresentados nas Tabelas A1 a A4, no Apêndice. De modo geral, os coeficientes estimados não contradisseram aos efeitos esperados, seja na média dos salários, ou ao longo dos quantis incondicionais. No entanto, para algumas dessas variáveis, os efeitos não são constantes ao longo da distribuição, o que justifica a análise por regressões quantílicas, além de diferirem entre as duas regiões.

Os coeficientes estimados para os níveis de escolaridade mostram que os retornos educacionais, apesar de serem estritamente positivos (tendo como base indivíduos que não concluiriam o 1º ciclo do ensino fundamental), não apresentaram um padrão muito bem definido até o nível de ensino médio completo. Os retornos do ensino médio e do ensino superior se destacam a partir do 5º decil, quando o retorno do ensino superior completo cresce de forma acentuada em relação aos decis da distribuição. Uma comparação entre 2002/2003 e 2012/2013, mostra a redução dos retornos educacionais ao longo do tempo, ao ponto em que nesse segundo biênio, os retornos são maiores na região Nordeste, provavelmente em consequência de uma relativa escassez de oferta de trabalhadores mais qualificados nessa região.

Com relação aos retornos à experiência (mensurados pelo efeito do grupo etário), verifica-se que são crescentes ao longo dos decis. Na comparação temporal, um comportamento diferenciado pode ser observado entre as regiões. Enquanto que nos decis mais altos, os retornos cresceram na região Nordeste, estes apresentaram uma redução no Sudeste. Já as variáveis relacionadas à discriminação (gênero e cor/raça) mostram um maior efeito discriminatório na região Nordeste, principalmente em relação às mulheres nas caudas da distribuição de rendimentos. E estes coeficientes não apresentaram mudanças significativas entre os dois períodos considerados.

[9] Para o caso específico das regressões para os decis, foi computada uma função de influência para cada observação da amostra usando estimativas amostrais de q_τ. Tais estimativas foram computadas a partir das densidades obtidas empregando a função de Kernel Epanechnikov e largura da banda de 0,06.

Com relação aos diferenciais setoriais, tendo como base o setor de comércio e serviços, não se observa um diferencial significativo para os ocupados no setor industrial. Já para o setor agrícola, o diferencial de rendimentos é negativo, mas decresce em termos absolutos quando se segue para os salários mais elevados. Destaca-se a magnitude da penalidade aos ocupados na agricultura nordestina, principalmente em 2012/2013.

Com relação ao tipo de ocupação, trabalhadores informais e por conta-própria apresentaram diferenciais negativos em relação aos formais, principalmente no Nordeste, e em termos absolutos esse diferencial diminui conforme se avança para decis mais elevados. No caso dos trabalhadores por conta-própria esse diferencial torna-se positivo na cauda superior da distribuição. No Sudeste os trabalhadores por conta-própria apresentaram relativa vantagem sobre os trabalhadores informais, no Nordeste essa vantagem surge apenas a partir do 4º decil.

Por fim, se observa diferenciais negativos para os residentes em zonas rurais (em relação a áreas urbanas) e positivos para os que residem em áreas metropolitanas (em relação aos residentes no interior). Os trabalhadores residentes em áreas rurais do Nordeste são mais penalizados na cauda inferior da distribuição de rendimentos do trabalho, enquanto que os prêmios salariais são maiores para os residentes nas áreas metropolitanas no Sudeste. Entre os dois períodos considerados, poucas mudanças foram observadas nesses diferenciais.

Por fim, o intercepto possui um peso crescente em relação aos quantis, em ambos os períodos e regiões.

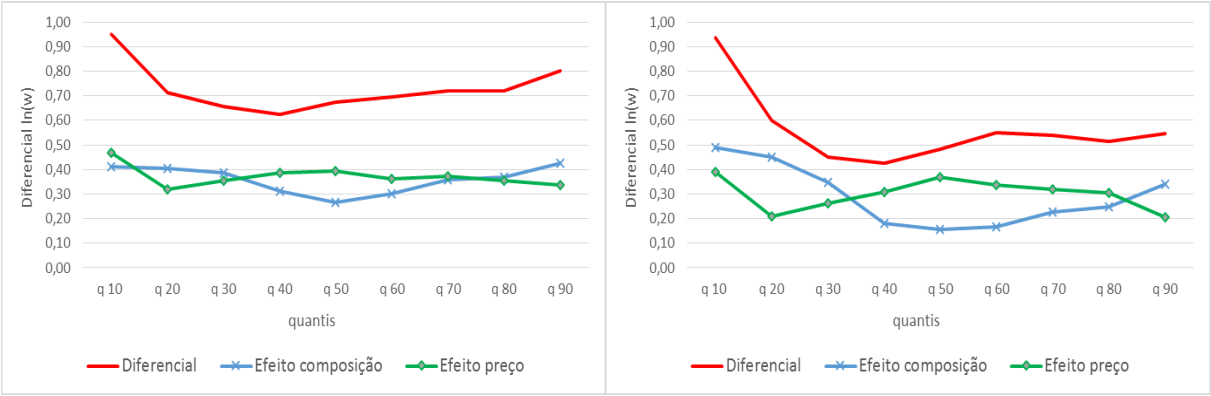
5.2. Decomposição dos diferenciais de salários

Uma análise das distribuições de características na seção anterior e das equações de rendimentos comentadas acima já fornecem evidências de como as duas regiões diferem não apenas em termos de atributos produtivos dos trabalhadores, mas também em termos da *estrutura salarial*.

Para decompor as diferenças de salários entre as regiões Sudeste e Nordeste, foi aplicado o método de decomposição com base em regressões RIF e o procedimento de construção de uma distribuição contrafactual pelo método de reponderação. A decomposição permite avaliar o quanto do diferencial de salários pode ser atribuído ao *efeito composição*, que capta as diferenças na distribuição de atributos produtivos, e o quanto pode ser explicado por diferenças nas estruturas salariais, ou seja, diferenças nos retornos dos atributos entre as duas regiões. O método também possibilita uma decomposição detalhada da contribuição de cada variável explicativa em termos dos dois efeitos, nos moldes da decomposição de Oaxaca-Blinder. Os gráficos apresentados nas Figuras 4, 5 e 6 apresentam os resultados da decomposição para o

diferencial regional em nove decis da distribuição de salários. Esses resultados também são sumarizados nas Tabelas A3 e A4, no Apêndice.

A Figura 4 mostra o diferencial de salários entre as duas regiões e destaca a decomposição desse diferencial em *efeito composição* e *estrutura salarial*. Ele mostra que ambos os efeitos são positivos, apontando que as diferenças salariais a favor dos indivíduos residentes na região Sudeste são determinadas pelos dois efeitos. Também pode-se verificar que a dominância de um efeito sobre o outro varia de acordo com diferentes pontos da distribuição de rendimentos, justificando a importância da análise quantílica.



2002/2003 **2012/2013**

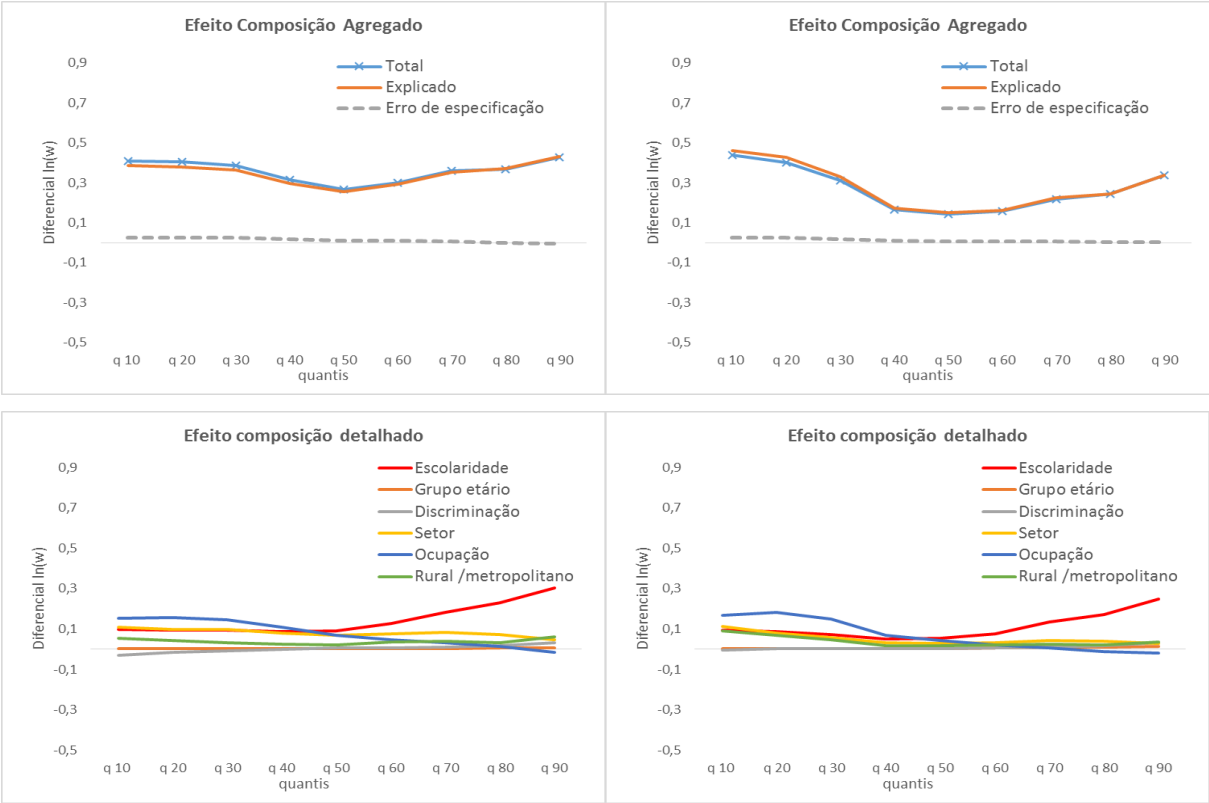
Figura 4. Decomposição do diferencial de salários Sudeste/Nordeste
Fonte: Estimações dos autores. Microdados do IBGE/PNAD.

Para 2002/2003, não se observa um padrão claro, com a predominância de um dos efeitos. Em 2012/2013 existe uma prevalência do *efeito composição* nos extremos da distribuição, principalmente na base da distribuição, indicando que nesse caso, as diferenças salariais são melhor explicadas por diferenças nas características dos trabalhadores. Em ambos os períodos, o efeito da *estrutura salarial* foi predominante na parte mais central da distribuição.

A Figura 5 apresenta gráficos que colocam em evidência o *efeito composição*. Os dois gráficos superiores apresentam o *efeito composição* "total" obtido com o procedimento de reponderação e o *efeito composição* "explicado". Em função do uso do procedimento de reponderação, em cada um desses gráficos também é apresentado o erro de especificação. Os valores desse termo de erro são muito próximos de zero, atestando a boa especificação do modelo e a consistência do *efeito composição* estimado pelo modelo reponderado (Firpo *et al.*, 2007).

De modo geral, o *efeito composição* mostra que as diferenças em termos de características dos trabalhadores contribuíram de forma significativa para o diferencial de

rendimentos, uma vez que o efeito é positivo ao longo de toda a distribuição. No caso dos diferenciais em 2012/2013, tem-se um formato de "U" mais acentuado, apontando novamente que as diferenças de características foram mais importantes na explicação do diferencial de rendimentos nos extremos da distribuição nesse biênio.



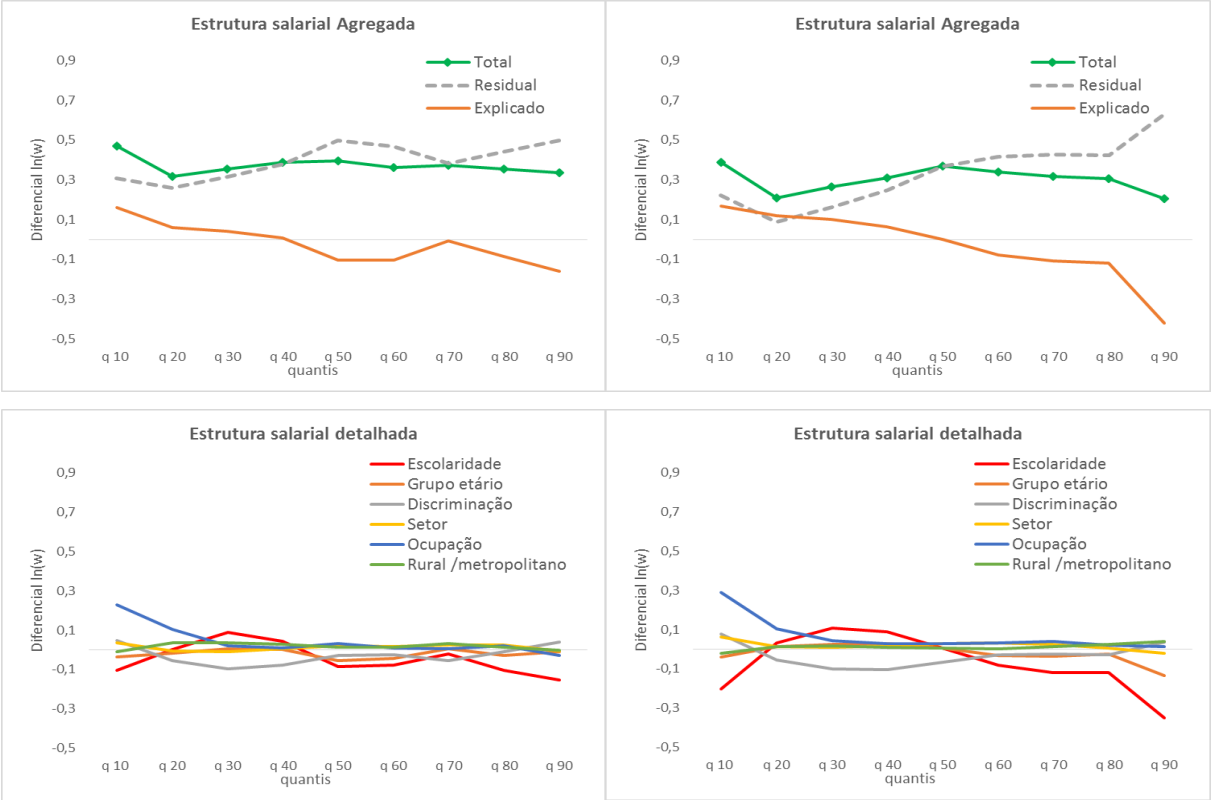
2002/2003 **2012/2013**

Figura 5. Decomposição do diferencial de salários Sudeste/Nordeste – *Efeito Composição*.
Fonte: Estimações dos autores. Microdados do IBGE/PNAD.

O detalhamento do *efeito composição*, apresentado nos dois gráficos inferiores, mostra que em ambos os períodos a escolaridade e a alocação ocupacional foram os fatores de maior importância. A baixa taxa de formalização dos trabalhadores no Nordeste em relação aos do Sudeste parece ser fundamental para explicar as diferenças de rendimentos nos quantis inferiores. Conforme avançamos para os quantis mais elevados, o efeito da formalização torna-se menos importante e a diferença na escolaridade passa a ser o principal determinante das diferenças salariais entre as regiões.

Por sua vez, o *efeito preço* é apresentado nos gráficos da Figura 6. Novamente, os dois gráficos superiores retratam os efeitos agregados das diferenças na *estrutura salarial* entre as

duas regiões. É evidente que as disparidades de retornos também são significantes para explicar as diferenças de rendimentos.



2002/2003 **2012/2013**

Figura 6. Decomposição do diferencial de salários Sudeste/Nordeste – *Estrutura salarial*.
Fonte: Estimações dos autores. Microdados do IBGE/PNAD.

Nesses gráficos também são destacados dois componentes do efeito da *estrutura salarial*: a parte explicada pelo modelo e a diferença residual, correspondente a diferença entre os interceptos. De modo geral, a *estrutura salarial* "total" não apresenta um padrão diferenciado entre os decis da distribuição. Já o efeito explicado se apresentou decrescente, enquanto que a diferença residual é crescente ao longo dos decis e de magnitude considerável. Esse comportamento do efeito explicado e do termo residual são mais acentuados no biênio 2012/2013.

Acerca do detalhamento da *estrutura salarial*, novamente os efeitos associados a ocupação e escolaridade se destacam. O *efeito preço* da ocupação, sendo positivo, revela uma contribuição a favor dos trabalhadores na região Sudeste. Contudo, esse efeito é especialmente importante nos primeiros decis da distribuição e, como pode ser visto nas regressões estimadas, decorre de maiores penalidades impostas aos trabalhadores informais e por conta própria no Nordeste. A contribuição do *efeito preço* associado à escolaridade variou entre os decis, mas

pode se dar destaque aos pontos em que ele foi negativo, favorável aos trabalhadores residentes no Nordeste. Esse comportamento pode ser relacionado ao diferencial positivo que retornos da educação no Nordeste apresentam em relação ao Sudeste. Como abordado anteriormente, espera-se retornos mais elevados onde existe uma relativa escassez de trabalhadores mais escolaridades.

6. CONSIDERAÇÕES FINAIS

O presente artigo analisou as diferenças nas distribuições de salários das regiões Nordeste e Sudeste do Brasil em dois biênios distintos, 2002/2003 e 2012/2013. Com inspiração na literatura que aponta que as diferenças regionais de renda são explicadas pelas disparidades na distribuição de atributos produtivos, principalmente educação, o estudo aplicou o método de decomposição com base em regressões RIF apresentado por Firpo *et al*. (2007) para avaliar os determinantes dos diferenciais de rendimentos do trabalho entre as regiões.

Dadas as possibilidades da metodologia adotada, os resultados encontrados mostram que as diferenças em termos de características individuais dos trabalhadores explicam em grande medida o diferencial de rendimentos entre as regiões analisadas. Esse resultado parece corroborar com a hipótese adotada nos trabalhos mais recentes, indicando que políticas com o objetivo de reduzir as diferenças educacionais (de capital humano de modo geral) seriam eficientes no combate às disparidades regionais de renda, especialmente nos estratos mais elevados da distribuição de rendimentos do trabalho. No entanto, os resultados permitem verificar que, garantir a equalização dos níveis de escolaridade da força de trabalho entre as regiões, não necessariamente seria uma ação suficiente para eliminar o diferencial de rendimentos.

Além das variáveis de capital humano, diferenças quanto à formalização apresentaram contribuições importantes para explicar o diferencial de rendimentos nos segmentos inferiores da distribuição. E as diferenças na *estrutura salarial* entre as regiões que, apesar de se apresentarem relativamente menos importantes, não podem ser ignoradas.

REFERÊNCIAS BIBLIOGRÁFICAS

BARSKY, R., BOUND, J., CHARLES, K. K., e LUPTON, J. P. Accounting for the black–white wealth gap: a nonparametric approach. ***Journal of the American Statistical Association***, v. 97, n. 459, p. 663-673, 2002.

BLACKABY, D. H.; MANNING, D. N. The North-South Divide: Questions of Existence and Stability?. ***Economic Journal***, v. 100, n. 401, p. 510-27, 1990.

BARROS, A. R. ***Desigualdades regionais no Brasil: natureza, causas, origens e soluções***. 2011.

BLINDER, A. Wage Discrimination: Reduced Form and Structural Estimates. ***Journal of Human Resources***, p. 436-455, 1973.

DINARDO, J; FORTIN, N. M.; LEMIEUX, T. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. ***Econometrica,*** 64: 1001-1044, 1996.

FARBER, S. C.; NEWMAN, R. J. Regional wage differentials and the spatial convergence of worker characteristic prices. ***The Review of Economics and Statistics***, p. 224-231, 1989.

FARBER, S. C.; NEWMAN, R. J. Accounting for South/non-South real wage differentials and for changes in those differentials over time. ***The Review of Economics and Statistics***, p. 215-223, 1987.

FIRPO, S.; FORTIN, N.; LEMIEUX, T.. Unconditional Quantile Regressions. ***Mimeo***. University of British Columbia e Pontifícia Universidade Católica do Rio de Janeiro,2006.

FIRPO, S.; FORTIN, N.; LEMIEUX, T.. Decomposing wage distributions using recentered influence function regressions. ***Mimeo***. University of British Columbia, 2007.

FIRPO, S.; FORTIN, N.; LEMIEUX, T.. Unconditional Quantile Regressions, ***Econometrica*** 77(3): 953-973, 2009.

FORTIN, N.; LEMIEUX, T.; FIRPO, S.. Decomposition methods in economics. ***Handbook of labor economics***, v. 4, p. 1-102, 2011.

FOURNIER, J.; KOSKE, I.. Less income inequality and more growth–Are they compatible? Part 7. The drivers of labour earnings inequality–An analysis based on conditional and unconditional quantile regressions. ***OECD Economics Department Working Papers***, n. 930, OECD Publishing, 2012.

GERKING, S. D.; WEIRICK, W. N.. Compensating differences and interregional wage differentials. ***The Review of Economics and Statistics***, p. 483-487, 1983.

HERRERA-IDÁRRAGA, P.; LÓPEZ-BAZO, E.; MOTELLÓN CORRAL, E. Regional wage gaps, education, and informality in an emerging country. The case of Colombia. ***IREA–Working Papers***, 2015, IR15/009, 2015.

KOENKER, R.. ***Quantile regression***. Cambridge university press, 2005.

KOENKER, R. e BASSETT JR, G.. Regression quantiles. ***Econometrica: journal of the Econometric Society***, p. 33-50, 1978.

LÓPEZ-BASO, E.; MOTELLÓN E.. Human capital and regional wage gaps. ***Regional Studies***, 46, 2012.

MACHADO, J. A. F.; MATA, J.. Counterfactual decomposition of changes in wage distributions using quantile regression. ***Journal of applied Econometrics***, v. 20, n. 4, p. 445-465, 2005.

MENEZES, T. A.; AZZONI, C. R. Convergência de salários entre as regiões metropolitanas brasileiras: custo de vida e aspectos de demanda e oferta de trabalho. ***Pesquisa e planejamento econômico***, v. 36, n. 3, p. 449-470, 2006.

MOTELLÓN, E.; LÓPEZ-BAZO, E.; EL-ATTAR, M.. Regional heterogeneity in wage distributions: evidence from Spain. ***Journal of Regional Science***, v. 51, n. 3, p. 558-584, 2011.

NEUMAN, S.; e OAXACA R. Wage decompositions with selectivity-corrected wage equations: A methodological note. ***The Journal of Economic Inequality***, v. 2, n. 1, p. 3-10, 2004.

OAXACA, R. L.. Male-Female Wage Diferentials in Urban Labor Markets. ***International Economic Review***, 14(3), pg 693-709, 1973.

OAXACA, R. L.; RANSOM, M. R. Identification in detailed wage decompositions. ***Review of Economics and Statistics***, v. 81, n. 1, p. 154-157, 1999.

PEREIRA J.; GALEGO A.. Inter-regional wage differentials in Portugal: An analysis across the wage distribution. ***Regional Studies***, 48, 2014.

PESSOA, S. Existe um problema de desigualdade regional no Brasil. ***Encontro Nacional de Economia***, v. 29, 2001.

SALVATO, M. A.; FERREIRA, P. C.; DUARTE, A. J. M.. O impacto da escolaridade sobre a distribuição de renda. **Estudos Econômicos** (São Paulo), v. 40, n. 4, p. 753-791, 2010.

SAVEDOFF, William D. Os diferenciais regionais de salários no Brasil: segmentação versus dinamismo da demanda. ***Pesquisa e Planejamento Econômico***, v. 20, n. 3, p. 521-556, 1990.

YUN, M.. A simple solution to the identification problem in detailed wage decompositions. ***Economic inquiry***, v. 43, n. 4, p. 766-772, 2005.

YUN, M.. Identification problem and detailed Oaxaca decomposition: a general solution and inference. ***Journal of economic and social measurement***, v. 33, n. 1, p. 27, 2008.

APÊNDICE

Tabela A1: Coeficientes das regressões RIF - Nordeste (2002/2003)

	MQO	***q10***	***q20***	***q30***	***q40***	***q50***	***q60***	***q70***	***q80***	***q90***
Fund. 1º ciclo completo	0,200***	0,248***	0,214***	0,198***	0,160***	0,139***	0,172***	0,209***	0,198***	0,168***
	(0,009)	(0,020)	(0,014)	(0,011)	(0,008)	(0,007)	(0,009)	(0,011)	(0,014)	(0,018)
Fundamental completo	0,386***	0,370***	0,337***	0,325***	0,279***	0,275***	0,348***	0,445***	0,450***	0,462***
	(0,012)	(0,028)	(0,019)	(0,015)	(0,010)	(0,009)	(0,012)	(0,015)	(0,019)	(0,025)
Médio completo	0,713***	0,442***	0,454***	0,460***	0,430***	0,465***	0,651***	0,923***	1,092***	1,193***
	(0,011)	(0,025)	(0,017)	(0,013)	(0,010)	(0,009)	(0,011)	(0,014)	(0,017)	(0,023)
Superior completo	1,723***	0,479***	0,517***	0,553***	0,540***	0,640***	1,026***	1,662***	2,441***	4,023***
	(0,021)	(0,049)	(0,034)	(0,026)	(0,019)	(0,017)	(0,021)	(0,027)	(0,033)	(0,045)
Idade 30-34 anos	0,096***	0,035	0,039*	0,051***	0,053***	0,073***	0,092***	0,136***	0,145***	0,175***
	(0,010)	(0,024)	(0,016)	(0,012)	(0,009)	(0,008)	(0,010)	(0,013)	(0,016)	(0,022)
Idade 35-39 anos	0,177***	0,100***	0,094***	0,098***	0,091***	0,108***	0,156***	0,229***	0,264***	0,305***
	(0,010)	(0,025)	(0,017)	(0,013)	(0,009)	(0,008)	(0,010)	(0,014)	(0,016)	(0,022)
Idade 40-44 anos	0,194***	0,145***	0,124***	0,113***	0,094***	0,118***	0,169***	0,246***	0,266***	0,320***
	(0,011)	(0,026)	(0,018)	(0,014)	(0,010)	(0,009)	(0,011)	(0,015)	(0,018)	(0,024)
Idade 45-49 anos	0,255***	0,163***	0,152***	0,145***	0,131***	0,156***	0,214***	0,322***	0,385***	0,457***
	(0,012)	(0,029)	(0,020)	(0,015)	(0,011)	(0,010)	(0,012)	(0,016)	(0,019)	(0,026)
Idade 50-54 anos	0,270***	0,168***	0,170***	0,177***	0,152***	0,175***	0,231***	0,331***	0,360***	0,445***
	(0,013)	(0,031)	(0,021)	(0,016)	(0,012)	(0,011)	(0,013)	(0,017)	(0,021)	(0,028)
Idade 55-65 anos	0,198***	0,059*	0,049*	0,099***	0,113***	0,135***	0,196***	0,291***	0,323***	0,377***
	(0,013)	(0,029)	(0,020)	(0,015)	(0,011)	(0,010)	(0,012)	(0,016)	(0,020)	(0,027)
Gênero (mulher=1)	-0,371***	-0,544***	-0,441***	-0,316***	-0,217***	-0,185***	-0,250***	-0,303***	-0,320***	-0,414***
	(0,007)	(0,017)	(0,012)	(0,009)	(0,006)	(0,006)	(0,007)	(0,009)	(0,011)	(0,016)
Cor (Negro | Pardo=1)	-0,072***	-0,003	-0,025*	-0,033***	-0,035***	-0,046***	-0,061***	-0,078***	-0,097***	-0,159***
	(0,007)	(0,017)	(0,012)	(0,009)	(0,006)	(0,006)	(0,007)	(0,009)	(0,011)	(0,016)
Agrícola	-0,427***	-0,587***	-0,477***	-0,448***	-0,351***	-0,318***	-0,371***	-0,418***	-0,405***	-0,313***
	(0,011)	(0,025)	(0,017)	(0,013)	(0,009)	(0,008)	(0,010)	(0,014)	(0,017)	(0,023)
Indústria	-0,062***	-0,136***	-0,029*	0,002	0,029***	0,013	-0,027**	-0,047***	-0,099***	-0,149***
	(0,009)	(0,020)	(0,014)	(0,011)	(0,008)	(0,007)	(0,009)	(0,011)	(0,014)	(0,019)
Informal	-0,299***	-0,300***	-0,449***	-0,524***	-0,449***	-0,332***	-0,306***	-0,282***	-0,208***	-0,113***
	(0,009)	(0,021)	(0,015)	(0,011)	(0,008)	(0,007)	(0,009)	(0,012)	(0,014)	(0,019)
Conta própria	-0,285***	-0,678***	-0,645***	-0,556***	-0,393***	-0,225***	-0,126***	-0,068***	0,008	0,121***
	(0,009)	(0,021)	(0,014)	(0,011)	(0,008)	(0,007)	(0,009)	(0,011)	(0,014)	(0,019)
Zona Rural	-0,051***	-0,120***	-0,068***	-0,062***	-0,041***	-0,018*	-0,041***	-0,041***	-0,028	-0,055**
	(0,009)	(0,022)	(0,015)	(0,011)	(0,008)	(0,007)	(0,009)	(0,012)	(0,015)	(0,020)
Área Metropolitana	0,113***	0,100***	0,100***	0,075***	0,055***	0,061***	0,094***	0,116***	0,103***	0,178***
	(0,008)	(0,020)	(0,014)	(0,010)	(0,007)	(0,007)	(0,008)	(0,011)	(0,013)	(0,018)
Intercepto	0,969***	0,294***	0,714***	0,918***	0,978***	0,949***	1,015***	1,104***	1,296***	1,609***
	(0,013)	(0,032)	(0,022)	(0,017)	(0,012)	(0,011)	(0,013)	(0,018)	(0,021)	(0,029)
N	*54080*	54111	54111	54111	54111	54111	54111	54111	54111	54111
r2	0,331	0,092	0,150	0,211	0,255	0,246	0,243	0,234	0,220	0,206
F	1483,088	304,651	530,650	805,278	1029,415	981,445	965,139	918,159	849,541	779,831

Fonte: Estimações dos autores/ microdados IBGE/PNAD.
Erros-padrão entre parênteses. * p < 0.05, ** p < 0.01, *** p < 0.001.

Tabela A2: Coeficientes das regressões RIF - Sudeste (2002/2003)

	MQO	*q10*	*q20*	*q30*	*q40*	*q50*	*q60*	*q70*	*q80*	*q90*
Fund. 1º ciclo completo	0,105***	0,114***	0,099***	0,111***	0,099***	0,091***	0,095***	0,080***	0,061***	0,031
	(0,010)	(0,010)	(0,008)	(0,009)	(0,010)	(0,010)	(0,011)	(0,014)	(0,016)	(0,025)
Fundamental completo	0,226***	0,208***	0,193***	0,212***	0,202***	0,202***	0,220***	0,210***	0,177***	0,160***
	(0,011)	(0,010)	(0,009)	(0,010)	(0,010)	(0,011)	(0,012)	(0,015)	(0,017)	(0,027)
Médio completo	0,415***	0,259***	0,283***	0,356***	0,381***	0,399***	0,446***	0,481***	0,460***	0,457***
	(0,010)	(0,010)	(0,008)	(0,009)	(0,010)	(0,010)	(0,011)	(0,014)	(0,016)	(0,025)
Superior completo	1,156***	0,336***	0,412***	0,588***	0,716***	0,854***	1,076***	1,395***	1,682***	2,432***
	(0,012)	(0,012)	(0,010)	(0,011)	(0,012)	(0,013)	(0,014)	(0,017)	(0,020)	(0,031)
Idade 30-34 anos	0,112***	0,018*	0,027***	0,057***	0,087***	0,106***	0,123***	0,142***	0,152***	0,198***
	(0,009)	(0,009)	(0,007)	(0,008)	(0,009)	(0,009)	(0,010)	(0,012)	(0,014)	(0,022)
Idade 35-39 anos	0,168***	0,035***	0,059***	0,087***	0,127***	0,155***	0,199***	0,225***	0,230***	0,274***
	(0,009)	(0,009)	(0,008)	(0,008)	(0,009)	(0,009)	(0,011)	(0,013)	(0,015)	(0,023)
Idade 40-44 anos	0,175***	0,042***	0,062***	0,100***	0,132***	0,154***	0,194***	0,225***	0,239***	0,301***
	(0,010)	(0,009)	(0,008)	(0,009)	(0,009)	(0,010)	(0,011)	(0,013)	(0,015)	(0,024)
Idade 45-49 anos	0,188***	0,028**	0,065***	0,101***	0,142***	0,166***	0,217***	0,249***	0,273***	0,376***
	(0,010)	(0,010)	(0,008)	(0,009)	(0,009)	(0,010)	(0,011)	(0,014)	(0,016)	(0,025)
Idade 50-54 anos	0,218***	0,028**	0,059***	0,107***	0,143***	0,186***	0,250***	0,299***	0,320***	0,444***
	(0,011)	(0,010)	(0,009)	(0,010)	(0,010)	(0,011)	(0,012)	(0,015)	(0,017)	(0,027)
Idade 55-65 anos	0,214***	0,019	0,058***	0,097***	0,141***	0,178***	0,234***	0,278***	0,329***	0,458***
	(0,010)	(0,010)	(0,008)	(0,009)	(0,010)	(0,011)	(0,012)	(0,014)	(0,016)	(0,026)
Gênero (mulher=1)	-0,295***	-0,148***	-0,179***	-0,255***	-0,299***	-0,317***	-0,318***	-0,333***	-0,302***	-0,323***
	(0,006)	(0,006)	(0,005)	(0,005)	(0,006)	(0,006)	(0,007)	(0,008)	(0,009)	(0,015)
Cor (Negro | Pardo=1)	-0,142***	-0,061***	-0,077***	-0,099***	-0,109***	-0,118***	-0,131***	-0,157***	-0,181***	-0,247***
	(0,005)	(0,005)	(0,004)	(0,005)	(0,005)	(0,006)	(0,006)	(0,008)	(0,009)	(0,014)
Agrícola	-0,197***	-0,256***	-0,241***	-0,237***	-0,226***	-0,207***	-0,204***	-0,181***	-0,117***	-0,041
	(0,013)	(0,012)	(0,010)	(0,012)	(0,012)	(0,013)	(0,015)	(0,018)	(0,020)	(0,032)
Indústria	0,051***	0,048***	0,051***	0,070***	0,067***	0,064***	0,055***	0,049***	0,027**	0,028
	(0,006)	(0,006)	(0,005)	(0,006)	(0,006)	(0,006)	(0,007)	(0,008)	(0,010)	(0,016)
Informal	-0,117***	-0,264***	-0,163***	-0,120***	-0,094***	-0,065***	-0,037***	-0,010	0,007	0,022
	(0,007)	(0,007)	(0,006)	(0,007)	(0,007)	(0,008)	(0,008)	(0,010)	(0,012)	(0,019)
Conta própria	0,005	-0,220***	-0,095***	-0,004	0,032***	0,077***	0,133***	0,197***	0,230***	0,209***
	(0,007)	(0,006)	(0,005)	(0,006)	(0,006)	(0,007)	(0,008)	(0,009)	(0,011)	(0,017)
Zona Rural	-0,164***	-0,193***	-0,141***	-0,152***	-0,146***	-0,136***	-0,111***	-0,112***	-0,092***	-0,103***
	(0,012)	(0,012)	(0,010)	(0,011)	(0,012)	(0,013)	(0,014)	(0,017)	(0,020)	(0,031)
Área Metropolitana	0,135***	0,056***	0,068***	0,081***	0,071***	0,082***	0,076***	0,091***	0,127***	0,232***
	(0,006)	(0,005)	(0,005)	(0,005)	(0,005)	(0,006)	(0,006)	(0,008)	(0,009)	(0,014)
Intercepto	1,672***	1,259***	1,345***	1,423***	1,513***	1,625***	1,677***	1,798***	1,989***	2,287***
	(0,013)	(0,012)	(0,010)	(0,012)	(0,012)	(0,013)	(0,014)	(0,018)	(0,020)	(0,032)
N	*60491*	60491	60491	60491	60491	60491	60491	60491	60491	60491
r2	0,267	0,123	0,144	0,164	0,179	0,191	0,2	0,206	0,21	0,178
F	1226,555	470,067	563,947	657,907	733,082	792,771	842,12	871,968	892,865	728,742

Fonte: Estimações dos autores/ microdados IBGE/PNAD.
Erros-padrão entre parênteses. * p < 0.05, ** p < 0.01, *** p < 0.001.

Tabela A3: Coeficientes das regressões RIF - Nordeste (2012/2013)

	MQO	*q10*	*q20*	*q30*	*q40*	*q50*	*q60*	*q70*	*q80*	*q90*
Fund. 1º ciclo completo	0,165***	0,309***	0,225***	0,157***	0,086***	0,081***	0,094***	0,139***	0,139***	0,149***
	(0,010)	(0,026)	(0,016)	(0,011)	(0,006)	(0,006)	(0,008)	(0,012)	(0,015)	(0,024)
Fundamental completo	0,283***	0,410***	0,368***	0,282***	0,153***	0,150***	0,186***	0,281***	0,299***	0,336***
	(0,012)	(0,031)	(0,019)	(0,013)	(0,007)	(0,007)	(0,009)	(0,014)	(0,018)	(0,029)
Médio completo	0,502***	0,586***	0,510***	0,418***	0,263***	0,280***	0,371***	0,590***	0,684***	0,790***
	(0,011)	(0,028)	(0,017)	(0,012)	(0,006)	(0,006)	(0,008)	(0,013)	(0,016)	(0,026)
Superior completo	1,278***	0,686***	0,676***	0,601***	0,439***	0,530***	0,802***	1,482***	2,083***	3,358***
	(0,017)	(0,045)	(0,028)	(0,019)	(0,010)	(0,010)	(0,013)	(0,021)	(0,026)	(0,042)
Idade 30-34 anos	0,080***	0,029	0,059**	0,040**	0,034***	0,058***	0,079***	0,135***	0,142***	0,177***
	(0,011)	(0,029)	(0,018)	(0,012)	(0,007)	(0,007)	(0,008)	(0,013)	(0,016)	(0,027)
Idade 35-39 anos	0,137***	0,076*	0,100***	0,087***	0,063***	0,091***	0,125***	0,204***	0,221***	0,276***
	(0,011)	(0,030)	(0,019)	(0,013)	(0,007)	(0,007)	(0,009)	(0,014)	(0,017)	(0,028)
Idade 40-44 anos	0,151***	0,112***	0,114***	0,095***	0,070***	0,098***	0,134***	0,230***	0,251***	0,293***
	(0,012)	(0,032)	(0,020)	(0,013)	(0,007)	(0,007)	(0,009)	(0,014)	(0,018)	(0,029)
Idade 45-49 anos	0,186***	0,132***	0,144***	0,102***	0,078***	0,098***	0,148***	0,254***	0,302***	0,358***
	(0,013)	(0,033)	(0,021)	(0,014)	(0,008)	(0,008)	(0,010)	(0,015)	(0,019)	(0,031)
Idade 50-54 anos	0,200***	0,100**	0,134***	0,095***	0,079***	0,114***	0,165***	0,285***	0,337***	0,436***
	(0,014)	(0,036)	(0,023)	(0,015)	(0,008)	(0,008)	(0,010)	(0,016)	(0,021)	(0,034)
Idade 55-65 anos	0,195***	-0,014	0,076***	0,083***	0,081***	0,122***	0,169***	0,307***	0,363***	0,556***
	(0,013)	(0,035)	(0,022)	(0,015)	(0,008)	(0,008)	(0,010)	(0,016)	(0,020)	(0,033)
Gênero (mulher=1)	-0,322***	-0,598***	-0,415***	-0,262***	-0,134***	-0,136***	-0,200***	-0,290***	-0,320***	-0,381***
	(0,008)	(0,020)	(0,012)	(0,008)	(0,005)	(0,005)	(0,006)	(0,009)	(0,011)	(0,019)
Cor (Negro | Pardo=1)	-0,068***	-0,060**	-0,066***	-0,045***	-0,027***	-0,028***	-0,042***	-0,077***	-0,109***	-0,170***
	(0,008)	(0,021)	(0,013)	(0,009)	(0,005)	(0,005)	(0,006)	(0,009)	(0,012)	(0,019)
Agrícola	-0,486***	-0,990***	-0,682***	-0,457***	-0,247***	-0,237***	-0,288***	-0,390***	-0,379***	-0,328***
	(0,012)	(0,031)	(0,019)	(0,013)	(0,007)	(0,007)	(0,009)	(0,014)	(0,018)	(0,029)
Indústria	-0,020*	-0,038	0,022	0,049***	0,037***	0,026***	-0,007	-0,036***	-0,064***	-0,097***
	(0,009)	(0,023)	(0,014)	(0,009)	(0,005)	(0,005)	(0,007)	(0,010)	(0,013)	(0,021)
Informal	-0,230***	-0,409***	-0,621***	-0,556***	-0,282***	-0,190***	-0,138***	-0,105***	-0,045***	-0,059**
	(0,009)	(0,024)	(0,015)	(0,010)	(0,006)	(0,005)	(0,007)	(0,011)	(0,014)	(0,022)
Conta própria	-0,267***	-0,878***	-0,793***	-0,597***	-0,265***	-0,144***	-0,042***	0,046***	0,140***	0,205***
	(0,009)	(0,023)	(0,014)	(0,010)	(0,005)	(0,005)	(0,007)	(0,010)	(0,013)	(0,021)
Zona Rural	-0,131***	-0,400***	-0,252***	-0,156***	-0,056***	-0,035***	-0,030***	-0,027*	-0,032*	-0,075**
	(0,010)	(0,027)	(0,017)	(0,011)	(0,006)	(0,006)	(0,008)	(0,012)	(0,015)	(0,025)
Área Metropolitana	0,099***	0,122***	0,132***	0,094***	0,041***	0,049***	0,069***	0,083***	0,066***	0,103***
	(0,008)	(0,022)	(0,014)	(0,009)	(0,005)	(0,005)	(0,006)	(0,010)	(0,013)	(0,020)
Intercepto	1,419***	0,921***	1,300***	1,458***	1,401***	1,403***	1,392***	1,438***	1,616***	1,904***
	(0,015)	(0,039)	(0,024)	(0,016)	(0,009)	(0,009)	(0,011)	(0,018)	(0,022)	(0,036)
N	*51841*	51841	51841	51841	51841	51841	51841	51841	51841	51841
r2	0,279	0,138	0,221	0,266	0,246	0,21	0,192	0,183	0,179	0,153
F	1113,09	461,889	817,25	1042,003	938,501	767,372	681,935	643,682	628,555	519,553

Fonte: Estimações dos autores/ microdados IBGE/PNAD.
Erros-padrão entre parênteses. * p < 0.05, ** p < 0.01, *** p < 0.001.

Tabela A4: Coeficientes das regressões RIF - Sudeste (2012/2013)

	MQO	*q10*	*q20*	*q30*	*q40*	*q50*	*q60*	*q70*	*q80*	*q90*
Fund. 1º ciclo completo	0,105***	0,114***	0,099***	0,111***	0,099***	0,091***	0,095***	0,080***	0,061***	0,031
	(0,010)	(0,010)	(0,008)	(0,009)	(0,010)	(0,010)	(0,011)	(0,014)	(0,016)	(0,025)
Fundamental completo	0,226***	0,208***	0,193***	0,212***	0,202***	0,202***	0,220***	0,210***	0,177***	0,160***
	(0,011)	(0,010)	(0,009)	(0,010)	(0,010)	(0,011)	(0,012)	(0,015)	(0,017)	(0,027)
Médio completo	0,415***	0,259***	0,283***	0,356***	0,381***	0,399***	0,446***	0,481***	0,460***	0,457***
	(0,010)	(0,010)	(0,008)	(0,009)	(0,010)	(0,010)	(0,011)	(0,014)	(0,016)	(0,025)
Superior completo	1,156***	0,336***	0,412***	0,588***	0,716***	0,854***	1,076***	1,395***	1,682***	2,432***
	(0,012)	(0,012)	(0,010)	(0,011)	(0,012)	(0,013)	(0,014)	(0,017)	(0,020)	(0,031)
Idade 30-34 anos	0,112***	0,018*	0,027***	0,057***	0,087***	0,106***	0,123***	0,142***	0,152***	0,198***
	(0,009)	(0,009)	(0,007)	(0,008)	(0,009)	(0,009)	(0,010)	(0,012)	(0,014)	(0,022)
Idade 35-39 anos	0,168***	0,035***	0,059***	0,087***	0,127***	0,155***	0,199***	0,225***	0,230***	0,274***
	(0,009)	(0,009)	(0,008)	(0,008)	(0,009)	(0,009)	(0,011)	(0,013)	(0,015)	(0,023)
Idade 40-44 anos	0,175***	0,042***	0,062***	0,100***	0,132***	0,154***	0,194***	0,225***	0,239***	0,301***
	(0,010)	(0,009)	(0,008)	(0,009)	(0,009)	(0,010)	(0,011)	(0,013)	(0,015)	(0,024)
Idade 45-49 anos	0,188***	0,028**	0,065***	0,101***	0,142***	0,166***	0,217***	0,249***	0,273***	0,376***
	(0,010)	(0,010)	(0,008)	(0,009)	(0,009)	(0,010)	(0,011)	(0,014)	(0,016)	(0,025)
Idade 50-54 anos	0,218***	0,028**	0,059***	0,107***	0,143***	0,186***	0,250***	0,299***	0,320***	0,444***
	(0,011)	(0,010)	(0,009)	(0,010)	(0,010)	(0,011)	(0,012)	(0,015)	(0,017)	(0,027)
Idade 55-65 anos	0,214***	0,019	0,058***	0,097***	0,141***	0,178***	0,234***	0,278***	0,329***	0,458***
	(0,010)	(0,010)	(0,008)	(0,009)	(0,010)	(0,011)	(0,012)	(0,014)	(0,016)	(0,026)
Gênero (mulher=1)	-0,295***	-0,148***	-0,179***	-0,255***	-0,299***	-0,317***	-0,318***	-0,333***	-0,302***	-0,323***
	(0,006)	(0,006)	(0,005)	(0,005)	(0,006)	(0,006)	(0,007)	(0,008)	(0,009)	(0,015)
Cor (Negro | Pardo=1)	-0,142***	-0,061***	-0,077***	-0,099***	-0,109***	-0,118***	-0,131***	-0,157***	-0,181***	-0,247***
	(0,005)	(0,005)	(0,004)	(0,005)	(0,005)	(0,006)	(0,006)	(0,008)	(0,009)	(0,014)
Agrícola	-0,197***	-0,256***	-0,241***	-0,237***	-0,226***	-0,207***	-0,204***	-0,181***	-0,117***	-0,041
	(0,013)	(0,012)	(0,010)	(0,012)	(0,012)	(0,013)	(0,015)	(0,018)	(0,020)	(0,032)
Indústria	0,051***	0,048***	0,051***	0,070***	0,067***	0,064***	0,055***	0,049***	0,027**	0,028
	(0,006)	(0,006)	(0,005)	(0,006)	(0,006)	(0,006)	(0,007)	(0,008)	(0,010)	(0,016)
Informal	-0,117***	-0,264***	-0,163***	-0,120***	-0,094***	-0,065***	-0,037***	-0,010	0,007	0,022
	(0,007)	(0,007)	(0,006)	(0,007)	(0,007)	(0,008)	(0,008)	(0,010)	(0,012)	(0,019)
Conta própria	0,005	-0,220***	-0,095***	-0,004	0,032***	0,077***	0,133***	0,197***	0,230***	0,209***
	(0,007)	(0,006)	(0,005)	(0,006)	(0,006)	(0,007)	(0,008)	(0,009)	(0,011)	(0,017)
Zona Rural	-0,164***	-0,193***	-0,141***	-0,152***	-0,146***	-0,136***	-0,111***	-0,112***	-0,092***	-0,103***
	(0,012)	(0,012)	(0,010)	(0,011)	(0,012)	(0,013)	(0,014)	(0,017)	(0,020)	(0,031)
Área Metropolitana	0,135***	0,056***	0,068***	0,081***	0,071***	0,082***	0,076***	0,091***	0,127***	0,232***
	(0,006)	(0,005)	(0,005)	(0,005)	(0,005)	(0,006)	(0,006)	(0,008)	(0,009)	(0,014)
Intercepto	1,672***	1,259***	1,345***	1,423***	1,513***	1,625***	1,677***	1,798***	1,989***	2,287***
	(0,013)	(0,012)	(0,010)	(0,012)	(0,012)	(0,013)	(0,014)	(0,018)	(0,020)	(0,032)
N	*60491*	60491	60491	60491	60491	60491	60491	60491	60491	60491
r2	0,267	0,123	0,144	0,164	0,179	0,191	0,2	0,206	0,21	0,178
F	1226,555	470,067	563,947	657,907	733,082	792,771	842,12	871,968	892,865	728,742

Fonte: Estimações dos autores/ microdados IBGE/PNAD.
Erros-padrão entre parênteses. * p < 0.05, ** p < 0.01, *** p < 0.001.

Tabela A5: Resultados da decomposição do diferencial regional de salários: Sudeste - Nordeste (2002/2003).

	MQO	***q 10***	***q 20***	***q 30***	***q 40***	***q 50***	***q 60***	***q 70***	***q 80***	***q 90***
ln(w) Sudeste	1,594	0,710	0,939	1,154	1,323	1,491	1,664	1,911	2,203	2,717
ln(w) Nordeste	0,838	-0,239	0,226	0,499	0,698	0,817	0,970	1,190	1,482	1,914
Diferencial	0,756	0,949	0,713	0,655	0,625	0,674	0,694	0,720	0,722	0,803
Efeito composição	0,364	0,410	0,404	0,387	0,314	0,265	0,301	0,359	0,369	0,426
Estrutura salarial (Efeito preço)	0,382	0,468	0,318	0,356	0,387	0,393	0,362	0,373	0,353	0,336
Efeito composição detalhado										
Escolaridade	0,167	0,097	0,094	0,094	0,085	0,090	0,128	0,184	0,229	0,304
Grupo etário	0,004	0,004	0,004	0,002	0,002	0,002	0,003	0,004	0,005	0,006
Discriminação	0,004	-0,031	-0,017	-0,007	0,000	0,005	0,007	0,010	0,015	0,032
Setor	0,083	0,108	0,098	0,096	0,079	0,070	0,076	0,083	0,074	0,047
Ocupação	0,076	0,154	0,158	0,145	0,108	0,067	0,046	0,032	0,013	-0,016
Rural /metropolitano	0,041	0,052	0,041	0,033	0,023	0,020	0,034	0,040	0,034	0,060
Efeito preço detalhado										
Escolaridade	0,222	-0,106	0,000	0,089	0,044	-0,087	-0,078	-0,019	-0,106	-0,155
Grupo etário	-0,023	-0,038	-0,017	0,005	0,002	-0,055	-0,044	0,004	-0,028	-0,009
Discriminação	-0,028	0,045	-0,056	-0,099	-0,080	-0,029	-0,024	-0,055	-0,011	0,039
Setor	0,012	0,037	-0,006	-0,011	0,004	0,021	0,016	0,026	0,024	-0,007
Ocupação	0,040	0,230	0,102	0,021	0,010	0,030	0,010	0,007	0,021	-0,029
Rural /metropolitano	0,111	-0,010	0,036	0,037	0,027	0,015	0,014	0,031	0,013	0,000
Intercepto	0,048	0,308	0,258	0,314	0,380	0,499	0,467	0,381	0,440	0,498

Fonte: Estimações dos autores/ microdados IBGE/PNAD.

Tabela A6: Resultados da decomposição do diferencial regional de salários: Sudeste - Nordeste (2012/2013).

	MQO	***q 10***	***q 20***	***q 30***	***q 40***	***q 50***	***q 60***	***q 70***	***q 80***	***q 90***
ln(w) Sudeste	1,997	1,296	1,453	1,610	1,734	1,898	2,040	2,237	2,483	2,922
ln(w) Nordeste	1,400	0,361	0,855	1,160	1,309	1,416	1,491	1,698	1,968	2,374
Diferencial	0,597	0,935	0,598	0,449	0,426	0,482	0,549	0,539	0,515	0,548
Efeito composição	0,290	0,488	0,451	0,349	0,182	0,156	0,167	0,228	0,247	0,339
Estrutura salarial (Efeito preço)	0,316	0,389	0,208	0,264	0,308	0,370	0,339	0,319	0,305	0,206
Efeito composição detalhado										
Escolaridade	0,115	0,095	0,087	0,073	0,049	0,055	0,077	0,133	0,172	0,247
Grupo etário	0,006	0,002	0,003	0,003	0,002	0,003	0,005	0,008	0,010	0,014
Discriminação	0,008	-0,003	0,004	0,003	0,003	0,003	0,005	0,011	0,019	0,033
Setor	0,054	0,111	0,081	0,057	0,032	0,030	0,033	0,042	0,038	0,029
Ocupação	0,064	0,168	0,182	0,148	0,070	0,042	0,022	0,006	-0,014	-0,021
Urbano/metropolitano	0,043	0,091	0,069	0,046	0,018	0,016	0,020	0,023	0,020	0,035
Efeito preço detalhado										
Escolaridade	0,140	-0,204	0,032	0,108	0,087	0,005	-0,082	-0,121	-0,121	-0,351
Grupo etário	-0,016	-0,039	0,014	0,024	0,027	0,004	-0,034	-0,036	-0,024	-0,135
Discriminação	-0,034	0,079	-0,054	-0,102	-0,104	-0,067	-0,028	-0,025	-0,027	0,034
Setor	0,027	0,063	0,014	0,011	0,015	0,027	0,032	0,023	0,005	-0,021
Ocupação	0,066	0,291	0,103	0,045	0,029	0,028	0,032	0,038	0,022	0,012
Urbano/metropolitano	0,035	-0,020	0,011	0,016	0,008	0,005	0,002	0,011	0,026	0,041
Intercepto	0,097	0,221	0,088	0,162	0,246	0,368	0,417	0,428	0,425	0,627

Fonte: Estimações dos autores/ microdados IBGE/PNAD.