

Mismatch nos Mercados de Trabalho Regionais Brasileiros: O Que Explica as Diferenças Regionais?

Premiado em 3º Lugar no XV Encontro Regional de Economia, realizado pelo Banco do Nordeste do Brasil e Anpec, em Fortaleza, em 19 e 20 de julho de 2010. Dois trabalhos obtiveram esta classificação.

RESUMO

A partir dos resultados recentemente apresentados por Santos (2002) e Diaz e Machado (2008) acerca dos níveis de *mismatch* (indivíduos com níveis educacionais diversos daqueles requeridos para suas ocupações) nos mercados de trabalhos regionais brasileiros e utilizando microdados da mais recente Pesquisa Nacional por Amostra de Domicílios (IBGE, 2007), o trabalho investiga os determinantes destes desajustamentos e, a partir da decomposição proposta por Fairlie (2003) para parametrizações não-lineares, os fatores que explicam a posição mais desfavorável da região Nordeste, principalmente no que diz respeito à condição de subescolarizados (*undereducated*). As evidências obtidas indicam que tanto fatores de oferta de trabalho (níveis de escolaridade, principalmente) como de demanda (ramos de atividade, por exemplo) afetam os níveis de *mismatch* dos mercados regionais de trabalho. Ambos os conjuntos de fatores também explicariam os maiores níveis de desajustamento presentes na região Nordeste quando comparada à região Sudeste do país.

PALAVRAS-CHAVE

Mercado de Trabalho. Sobre-educação. Subeducação. Diferenças Regionais.

Moisés Freitas Athayde Cavalcanti

- Mestre em Economia pela Universidade Federal de Pernambuco (UFPE);
- Professor de Economia do Departamento de Letras e Ciências Humanas (DLCH)-UFPE.

Felipe Mota Campos

- Mestre e Doutorando em Economia pelo Programa de Pós-Graduação em Economia Pimes - UFPE.

Raul da Mota Silveira Neto

- Doutor em Economia pela Universidade de São Paulo (USP);
- Professor do Departamento de Economia e do Pimes-UFPE e Pesquisador do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

1– INTRODUÇÃO

A literatura econômica tem estudado a diferença existente entre a educação necessária para o desempenho de um determinado trabalho e a efetiva educação adquirida pelo trabalhador. Neste sentido, considera-se como sobre-educação o fenômeno em que o indivíduo apresenta nível educacional superior a aquele requerido pelo cargo que ocupa, incorrendo em menor retorno salarial quando comparado a indivíduos que trabalham em atividades que requerem o seu nível de formação. Sub-educação, em sentido contrário, indica o trabalhador que exerce atividade na qual se demanda um nível de qualificação superior a que efetivamente possui, gerando ganhos salariais superiores aos indivíduos que possuem o mesmo nível educacional, mas que trabalham em atividades que requerem exatamente a educação da qual dispõem. Os indivíduos que possuem nível de escolaridade compatível com a requerida pela sua ocupação são classificados como corretamente alocados.

A existência destes fenômenos pode sugerir um desalinhamento entre os sistemas de educação, público e privado, e a demanda efetiva do mercado de trabalho, gerando recursos humanos com perfis aquém ou além do demandado pelos diversos setores da economia. Na hipótese da existência da sobre-educação, este fenômeno pode indicar retornos abaixo do que é socialmente esperado, devido à realidade do mercado de trabalho, promovendo, em alguns casos, a desmotivação de trabalhadores em melhorar sua capacitação uma vez que o investimento em educação não irá retornar o valor esperado na forma de salário.

Em linha com as evidências derivadas de estudos para diferentes conjuntos de países, (GROOT; MAASEN VAN DEN BRINK, 2000), recentemente, Santos (2002) e Diaz e Machado (2008) mostraram que tais fenômenos se fazem presentes de forma importante nos mercados regionais brasileiros. Diaz e Machado (2008), em particular, indicaram que cerca de 70% do pessoal ocupado brasileiro apresenta algum tipo de *mismatch*, estando, sobretudo, na condição de subeducado (53%), e que a região Nordeste se apresenta na pior das situações entre as macrorregiões brasileiras, com cerca de 58% de seu pessoal ocupado

com níveis de escolaridade abaixo daquele exigido pelas ocupações disponíveis na região. Note-se que, do ponto de vista regional, tal evidência desfavorável ao Nordeste é preocupante, uma vez que, como mostrado por Santos (2002) e Diaz e Machado (2008), há uma penalidade de renda para o trabalhador associada à condição de subeducado.

Não obstante o mérito destes estudos iniciais, duas importantes lacunas se fazem presentes, para o caso brasileiro, a respeito da presença e dos níveis de *mismatch* dos mercados de trabalho regionais do país. Em primeiro lugar, os trabalhos existentes pouco discutem sobre os fatores determinantes ou associados aos níveis de *mismatch* presentes no país e em suas macrorregiões. Evidentemente, tais desajustamentos nos mercados de trabalho regionais devem estar vinculados a fatores de oferta (escolaridade, gênero etc.) e demanda (ramos de atividade, grau de formalização etc.) dos mercados de trabalho regionais, entretanto, muito pouco é explorado a respeito. Em segundo lugar, o grau de desconhecimento ainda é maior a respeito dos fatores explicativos do atraso da região Nordeste do país. Por exemplo, ainda que se possa, sugestivamente, especular a respeito da importância dos baixos níveis de escolaridade da população da referida região para explicar seus níveis de *mismatch* (principalmente para a situação de subeducação), deve-se considerar sua importância conjuntamente ao lado de fatores de demanda, como, por exemplo, a presença e peso na região de diferentes ramos de atividades.

Sob tal perspectiva, dois são os objetivos deste trabalho. De início, fazendo uso de um modelo simples de variável dependente binária (modelo *probit*), o trabalho pretende evidenciar as diferentes variáveis observáveis associadas às condições de *mismatch* (subeducado, sobre-educado) e à condição de ajustamento adequado do trabalhador brasileiro, enfatizando particularmente o papel dos fatores de oferta e demanda de trabalho nas evidências geradas. Em seguida, a partir de uma decomposição recentemente proposta por Fairlie (2003) aplicada a estimativas paramétricas não-lineares e no sentido de entender as razões por trás da precária situação nordestina, a pesquisa procura explicar o papel

das diferentes variáveis observáveis de demanda e oferta de trabalho na determinação das diferenças de *mismatch* entre a região Nordeste (mais pobre do país) e a região Sudeste (mais rica do país).

A partir de uma amostra de 408 famílias¹ ocupacionais da CBO-Domiciliar entre 511 existentes na Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2007, os resultados obtidos inicialmente indicam que cerca de 22,8% dos brasileiros são sobre-educados, com as regiões Nordeste e Sudeste do país apresentando percentuais, respectivamente, de 19,6% e 25%. Por outro lado, para a condição de subeducado, o Nordeste apresenta a maior taxa do país (32,8%), bem mais elevada que aquelas encontradas para o país (27,7%) e para o Sudeste (25%). Por sua vez, notou-se que tanto fatores de oferta de trabalho como de demanda estão associados às diferentes situações de *mismatch* dos mercados de trabalho regionais e que são sobretudo os fatores de oferta de trabalho que explicam a mais precária situação da região Nordeste.

Além desta seção, o trabalho está estruturado em mais cinco seções. A seguinte discorre sobre os elementos conceituais dos fenômenos e uma breve revisão dos principais trabalhos internacionais e nacionais e suas evidências. Na terceira parte, são apresentados e explicados os instrumentos metodológicos, a base de dados e variáveis utilizadas nos modelos, além das evidências acerca da sobre-educação, subeducação e indivíduos adequadamente alocados no Brasil e grandes regiões, a partir do modelo *Job Analysis*. Na quarta seção, são investigados os determinantes para condições de adequação e inadequação entre escolaridade requerida e observada dentro do mercado de trabalho brasileiro e seus efeitos marginais. A quinta parte dedica-se à análise da decomposição dos fatores que explicam as diferenças regionais entre Sudeste e Nordeste. Na conclusão, são reunidas as principais evidências obtidas para o *mismatch* no mercado de trabalho brasileiro e a contribuição das características individuais sobre o diferencial entre as regiões Nordeste e Sudeste.

¹ O termo família ocupacional aqui adotado representa um agrupamento de situações de emprego ou ocupações similares a quatro dígitos (para maiores informações vide Informações Gerais CBO, item 7). (BRASIL, 2010).

2 – SOBRE-EDUCAÇÃO E SUBEDUCAÇÃO NA TEORIA ECONÔMICA

A literatura voltada à incompatibilidade entre o grau de escolaridade efetiva e a requerida dos trabalhadores para suas práticas profissionais vem crescendo, nas últimas duas décadas, especialmente nos países de economia desenvolvida. Fundamentalmente, estes estudos se voltam em duas direções: análise das mudanças no retorno para a educação através do tempo e o desalinhamento entre a formação requerida pelo trabalho que ele executa na organização e o nível de educação efetivo do trabalhador.

Rumberger (1981) argumenta que trabalhadores na condição de sobre-educação indicam desequilíbrios ou ineficiências no mercado de trabalho e/ou no sistema educacional, enquanto outros destacam que estes indivíduos são subutilizados. Tendo em vista a teoria do capital humano e o ciclo de vida do trabalho, pode-se questionar se esta divergência entre grau de formação requerido e o efetivo seja um fenômeno de longo prazo, normalmente observado em certos estágios deste ciclo, ou os trabalhadores podem temporariamente trabalhar em ocupações que lhes proporcionam habilidades para serem usadas mais tarde em outro emprego, de nível mais alto e, provavelmente, de melhor remuneração, indicando um ótimo *ex-ante*, pois a pessoa desprenderia um limitado período de tempo em uma atividade para a qual fosse sobre-educado a fim de desenvolver experiência (*learning by doing*), habilitando-se a preencher, no momento seguinte, uma vaga em outra atividade/emprego, para a qual ele tenha a educação demandada.

Ducan e Hoffaman (1981) afirmam que, se a suposição implícita na teoria do capital humano é a de que as firmas e trabalhadores adaptam a qualificação exigida para o cargo e os investimentos educacionais, concomitantemente, e considerando mudanças na demanda e oferta no mercado de trabalho, logo, a ocorrência da sobre-educação torna-se um acontecimento de curto prazo, decorrente da ausência de alinhamento entre os trabalhadores e as empresas. Enquanto para Groot (1996), as evidências mostram que estes trabalhadores são relativamente improdutivos, à medida que ganham salários menores

quando comparados aos corretamente alocados, destacando ainda que, ao passar do tempo, a firma passa a dispor de maior número de informações quanto à produtividade do trabalhador durante a realização efetiva de suas obrigações.

Para estimar quais fatores influenciam a transição de indivíduos graduados a um trabalho, Dolton e Vignoles (2000) estimaram a duração da sobre-educação após término do curso superior. Os resultados indicaram que a maioria dos graduados que eram sobre-educados no primeiro emprego, após a graduação, permaneceram no mesmo trabalho seis anos após a conclusão da graduação, levantando a hipótese de que a sobre-educação pode ser um fenômeno permanente para trabalhadores que se mantêm em atividades para as quais são sobre-educados, tornando compatível com a perspectiva de que trabalhadores sobre-educados teriam menor capacidade, se comparados a outros indivíduos com nível de educação idêntico, mas que exercem trabalhos que requerem a formação adquirida. Esta hipótese contradiz a ideia de que as pessoas podem ser temporariamente sobre-educadas devido a um *matching* ruim ou porque elas estão substituindo educação extra por outras formas de capital humano.

Para Bauer (2002), o controle da heterogeneidade não-observada pode ser relevante na hipótese da probabilidade de *mismatch* educacional estar correlacionada com as habilidades naturais dos indivíduos. Desta forma, uma grande dificuldade da literatura decorre do fato de que a quantidade de evidências empíricas disponíveis sobre o assunto é desenvolvida através de análises *cross-section*. Para comprovar sua assertiva, o autor, utilizando-se de grande painel de empregados da economia alemã, no período de 1984 a 1998, demonstra que os retornos sobre a *sobre-educação* são fortemente reduzidos e, geralmente, perdem significância estatística após o controle da heterogeneidade não-observada dos trabalhadores.

Colaborando com a teoria de sinalização proposta por Spence (1973), Cahuc e Zylberberg (2004) formalizam e desenvolvem um modelo explicativo para a condição de sobre-educação e utilizam algumas teorias para sua análise. Segundo estes autores, os

trabalhadores sob esta categorização têm, na condição de sobre-educação, um efeito substitutivo da habilidade ou produtividade que os indivíduos adequadamente alocados apresentam às firmas, embora não seja possível afirmar que a escolaridade aumente a produtividade do trabalhador, bem como a existência de correlação positiva entre duração dos estudos e o aumento dos rendimentos não prova a existência de um nexo de causalidade entre estas variáveis.

No caso brasileiro, para Santos (2002) em seu estudo sobre a presença da sobre-educação no mercado de trabalho no país, as empresas estariam contratando trabalhadores com uma formação educacional maior para poder treiná-los internamente, apresentando a hipótese de que se trata de uma consequência do baixo nível de formação geral e básica dos brasileiros, como resultado da ineficiência alocativa do Estado para os diversos níveis de educação. Com outro propósito, Machado; Oliveira e Carvalho (2003) identificam vários agrupamentos de ocupação que apresentam sobre-educação no Brasil, considerando-se sexo, idade, região e setor de atividade, e propõem uma tipologia ocupacional baseando-se na qualificação do trabalhador.

Mais recentemente, Diaz e Machado (2008) acrescentam novas evidências à incidência de *mismatch* no Brasil e suas grandes regiões, além de seus retornos, utilizando-se dos dados do Censo de 2000 e da Classificação Brasileira das Ocupações de 2002 (CBO). Seu trabalho é pioneiro no país, ao resolver o problema típico (endogeneidade) dos trabalhos voltados ao problema, utilizando a escolaridade requerida na CBO nas diversas famílias ocupacionais como elemento comparativo frente aos números observados na população brasileira e, ao regionalizar os resultados, permite comparar os diferentes níveis de *match* entre regiões com características próprias na dinâmica econômica e no mercado de trabalho.

Em seu estudo, contudo, os autores fundamentam sua metodologia de seleção das famílias ocupacionais a serem analisadas considerando a existência de um número significativo delas (508). Desta feita, separam os dados do Estado de São Paulo e calculam a frequência média dentro de cada grande grupo,

mantendo apenas aquelas que apresentaram número de ocupados acima da média observada, obtendo como resultado apenas 103 famílias ocupacionais. Tal procedimento poderia viesar os resultados, ao assumir o mercado de trabalho do Estado de São Paulo como padrão de análise para todo o Brasil e suas grandes regiões, o que pode desconsiderar por completo as diferenças econômicas regionais. Em seus resultados, o Brasil apresenta índice de subeducação de aproximadamente 53%; sobre-educados são 17,3% e adequados, aproximadamente, 29%. A região Sul apresentou a maior taxa de sobre-educação e a Sudeste a maior adequação entre as regiões, enquanto o Nordeste, a maior incidência de subeducação entre todos, com taxas ultrapassando os 58%. Em sua própria análise, os autores reconhecem o desalinhamento dos resultados encontrados para a região Nordeste frente ao trabalho do McGuinness (2006), supondo apenas ser reflexo do baixo nível de escolaridade da região, quando comparada com as regiões mais desenvolvidas do país, por suas palavras.

3 – EVIDÊNCIAS DO *MISMATCH* NO BRASIL: METODOLOGIA E BANCO DE DADOS

3.1 – Descrição da Base de Dados

A estimação da incidência da sobre-educação e da subeducação não é uma tarefa simples e certamente possui muitos pontos controversos dentro da literatura a este fenômeno dedicada. Fundamentalmente, existem três métodos diferentes para medir estes fenômenos: (i) *Job Analysis* (JA), (ii) *Realized Matches* (RM) e (iii) *Worker Self-Assessment* (WA). O primeiro (JA) consiste na avaliação sistemática baseada em uma determinada classificação ocupacional, na qual são definidas por analistas de trabalho todas as especificações dos tipos requeridos de instrução para os títulos do trabalho – método utilizado por Rumberger (1981); Hartog (2000) e Diaz e Machado (2008). O método *Realized Matches* estabelece que a instrução requerida é derivada a partir da média e do desvio-padrão de cada ocupação, onde o indivíduo é categorizado como inadequado se ele possui um nível de anos de estudo maior (menor) que a média mais (menos) o desvio-padrão de sua ocupação – utilizada

por Verdugo e Verdugo (1989) e Groot e Maassen Van Den Brink (2000). No método *Worker Self-Assessment*, o trabalhador especifica a qualificação requerida para o trabalho que desempenha – utilizado por Ducan e Hoffman (1981); Hartog e Tsang (1989) e Sicherman (1991).

Para os objetivos desse trabalho, utilizaremos o modelo JA para categorizar os indivíduos na condição de sobre-educado, subeducado ou de adequação. Para tanto, utilizamos a escolaridade requerida constante na Classificação Brasileira de Ocupações – CBO 2002, ajustada para a CBO-Domiciliar, que é uma adaptação desta adotada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para as pesquisas domiciliares, ajustada às padronizações nacional e internacional de classificação de ocupações.

As informações utilizadas no trabalho derivam dos microdados da Pesquisa Nacional por Amostra de Domicílios (PNAD) para o ano de 2007, do Instituto Brasileiro de Geografia e Estatística (IBGE). Tendo em mente que a inadequação entre os níveis de educação requeridos e os acumulados aplica-se somente a empregados ativos no mercado de trabalho, militares, trabalhadores não-remunerados, trabalhadores na produção para próprio consumo e trabalhadores na construção para próprio uso e afins não compõem a amostra utilizada. A pesquisa está focada nas pessoas de mais de dezoito e menos de sessenta e cinco anos, no período de referência, que residiam nas áreas urbanas das regiões metropolitanas brasileiras, com destaque para as regiões Nordeste e Sudeste do país, perfazendo um universo amostral total de 149.694 observações. Destes, 115.182 (76,9%) indivíduos estavam ocupados no período de referência em famílias ocupacionais que possuíam correspondência na CBO-2002 e que detinham requisitos de formação e experiência determinados. Permitindo-nos, portanto, tipificá-los quanto à qualidade do *match* entre seu montante de educação formal e as exigências de formação de cada família ocupacional.

Foram excluídas as ocupações dos grandes grupos 0 (Membros das forças armadas e auxiliares) e 1 (Dirigentes em geral), uma vez que, em sua maioria, não foi possível determinar a escolaridade requerida, representando 28 famílias ocupacionais destes dois

grandes grupos. Observaram-se e excluíram-se 54 famílias ocupacionais presentes na CBO domiciliar cujo código e/ou descrição não correspondiam perfeitamente à CBO 2002, evitando-se realizar qualquer tipo de análise pessoal e subjetiva com o objetivo de identificar equivalências entre códigos e descrições distintas. Da mesma forma, foram excluídas 21 famílias ocupacionais por não haver qualquer exigência de escolaridades ou requerer qualquer nível de escolaridade, tornando, pois, impossível tipificar os trabalhadores destas posições. Por fim, a amostra utilizada preservou 408 dentre as 511 famílias ocupacionais que a CBO domiciliar originalmente possui. Procurou-se, assim, evitar qualquer tipo de seleção *ad hoc* das famílias ocupacionais, tal como procedimentos adotados por Diaz e Machado (2008).²

Para determinação da escolaridade mínima e máxima requerida para as 408 famílias ocupacionais,

procedeu-se a uma análise das recomendações contidas na ficha de descrição das famílias, na seção “Formação e experiência”, onde se encontra descrita a escolaridade necessária, convertendo-as em montante de anos de estudo. Seguindo a mesma metodologia de Diaz e Machado (2008), para famílias ocupacionais que possuíam ocupações com níveis requeridos de formação diferentes, foram criados intervalos de anos de estudo para determinar a adequação dos trabalhadores destas famílias. Assim, o menor nível de formação requerido para uma ocupação dentro de cada família foi convertido em limite inferior, enquanto o maior nível de formação requerido para uma ocupação desta mesma família foi convertido em limite superior do intervalo de anos de estudos requeridos. Logo, indivíduos que possuíam um total de anos de estudo inferior ao limite mínimo exigido para sua família ocupacional foram tipificados como subeducados, ao

Tabela 1– Número de Famílias Ocupacionais Seleccionadas por Grandes Grupos

Grande Grupo	Composição dos Grupamentos Ocupacionais	Número de famílias ocupacionais	Número de famílias seleccionadas
0	Membros das forças armadas e auxiliares	15	0
1	Dirigentes em geral	13	0
2	Profissionais das ciências e das artes	77	65
3	Técnicos de nível médio	114	96
4	Trabalhadores de serviços administrativos	22	17
5	Trabalhadores dos serviços	36	28
6	Trabalhadores agrícolas	13	7
7	Trabalhadores da produção de bens e serviços industriais e de reparação e manutenção	134	125
8	Trabalhadores das indústrias de processos contínuos e outras indústrias	53	38
9	Trabalhadores de reparação e manutenção mecânica	34	32
	Total	511	408

Fonte: Elaboração Própria dos Autores.

² Em seu estudo, Diaz e Machado (2008) utilizaram os dados de pessoas do Estado de São Paulo para calcular o número médio de empregados dentro de cada grande grupo. Todas as famílias que apresentaram número de empregados acima desta média foram seleccionadas para fazer parte da base de dados. Este procedimento preservou apenas 103 famílias ocupacionais. Tal procedimento pode representar um viés de seleção, visto que a distribuição da força de trabalho do Estado de São Paulo não representa a dinâmica econômica e do mercado de trabalho dos outros estados da federação, além do Distrito Federal.

passo que aqueles que possuíam um total de anos de estudo superior ao limite máximo determinado foram classificados como sobre-educados.

Ademais, na Tabela 2, podemos constatar que, à exceção do grande grupo ocupacional 6 (Trabalhadores agrícolas), no qual a grande parte dos trabalhadores concentra-se em familiares ocupacionais, para as quais não há níveis de qualificação requeridos, a amostra utilizada parece conservar a estrutura dos grupos ocupacionais na PNAD-2007. Por exemplo, o grande grupo 2 (Profissionais das ciências e das artes) é composto por 12.841 (6,8%) indivíduos na PNAD-2007; destes, 10.585 estavam em famílias ocupacionais que tinham requisitos ocupacionais determinados e foram preservados na amostra utilizada, representando (10,7%). Assim, a amostra utilizada parece não trazer viés à distribuição dos indivíduos entre os grandes grupos ocupacionais observados na PNAD-2007. Desta feita, acredita-se que a amostra utilizada preserva a distribuição das características observáveis dos indivíduos nos mercados de trabalho regionais, permitindo-nos inferir sobre a importância destas e suas contribuições às incidências de adequação, sub e sobre-educação.

3.2 – Decomposição não-linear para Modelos de Reposta Binária: a Proposta de Fairlie (2003)

O modelo Oaxaca-Blinder é uma técnica comumente útil para identificar e quantificar separadamente as contribuições de diferentes características, tais como estado civil, escolaridade, experiência, localização geográfica, raça e gênero, dentre outros, para modelos estimados a partir de regressões lineares. Contudo, quando a variável dependente possui uma distribuição binária (por exemplo, participação do mercado) e a estimação de seus determinantes é feita com base em um modelo não-linear (*logit* ou *probit*), os coeficientes estimados não podem ser empregados diretamente na decomposição de Oaxaca-Blinder.

O primeiro modelo de decomposição para aplicação em estimações *logit* ou *probit* foi descrito por Fairlie (1999) para analisar as diferenças de empregabilidade entre negros/brancos. Neste artigo, o autor descreve uma técnica relativamente simples de identificar a importância de características dos indivíduos sobre o diferencial observado entre grupos/regiões de variáveis binárias.

Tabela 2 – Frequência Amostral por Grandes Grupos Ocupacionais na PNAD-2007 e na Amostra Utilizada

Grande Grupo Ocupacional	PNAD 2007		Amostra utilizada	
	Frequência	Percentual	Frequência	Percentual
0	1.443	0.76	0	0.00
1	9.212	4.87	0	0.00
2	12.841	6.79	10.585	9.19
3	14.473	7.66	12.342	10.72
4	16.228	8.59	13.200	11.46
5	58.463	30.93	42.585	36.97
6	32.851	17.38	2.266	1.97
7	35.069	18.55	28.966	25.15
8	4.461	2.36	1.694	1.47
9	3.983	2.11	3.544	3.08
Total	189.024	100	115.182	100

Fonte: Elaboração Própria dos Autores a partir dos Microdados da PNAD de 2007.

Aplicando a metodologia de Fairlie (1999, 2003), como primeiro passo, estimamos, através de um modelo *probit*, a importância de cada característica observável para o *mismatch* no mercado de trabalho (subeducado, adequado e sobre-educado).³ Assim, assumindo que os determinantes para o indivíduo (*i*) ser classificado como sobre-educado no mercado de trabalho (y_i), no período analisado, não são correlacionados ao vetor de erros e que $\varepsilon \sim (0, \sigma^2)$, através de um modelo *probit* padrão, estimamos a seguinte equação:

$$y_{ic}^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \quad (1)$$

em cuja equação (1) tem-se $y_{ic}^* = 1$, se o indivíduo *i* for categorizado como sobre-educado no período de referência, e igual a zero no caso contrário. A condição de *mismatch* é investigada em relação às características individuais do indivíduo *i*. O vetor x_{1i} capta o nível de capital humano detido pelo indivíduo, ou seja, *dummies* para o maior nível de educação formal atingido, experiência no mercado de trabalho e experiência ao quadrado, cuja influência é apreendida pelo vetor β_1 ($k \times 1$) de parâmetros desconhecidos. O vetor x_{2i} representa características familiares e demográficas, cor/raça declarada, gênero, estado civil, presença de filhos menores de 06 anos e de entre 07 e 14 anos, se reside em uma região metropolitana, se reside em zona rural, se migrou nos últimos quatro anos e em qual macrorregião reside atualmente (quando aplicável). O vetor β_2 é um vetor ($k \times 1$) de parâmetros desconhecidos relacionados às características individuais e demográficas que determinam a probabilidade de o indivíduo ser tipificado como sobre-educado. O vetor x_{3i} capta as características do modo de inserção do indivíduo no mercado de trabalho local, especificamente, se é sindicalizado, o tipo de ocupação e o ramo de atividade que exerce. Os valores estimados para os parâmetros do vetor β_3 permitirão estimarmos os efeitos dos

3 Por simplicidade, na metodologia exposta, as equações e modelos descritos referem-se apenas à investigação dos determinantes da sobre-educação e as contribuições destas para o diferencial observado entre regiões. Contudo, os determinantes para adequação e subeducação foram estimados seguindo a mesma metodologia e os resultados estimados são expostos nas seções seguintes.

mercados regionais sobre o diferencial de *mismatch* verificado entre as regiões Sudeste e Nordeste no período.

Seguindo Fairlie (2003), a decomposição para uma equação não-linear, tal como, $Y = F(X\hat{\beta})$, pode ser escrita como:

$$\bar{Y}_{se} - \bar{Y}_{ne} = \left[\sum_{i=1}^{N_{se}} \frac{F(X_i^{se} \hat{\beta}^{se})}{N_{se}} - \sum_{i=1}^{N_{ne}} \frac{F(X_i^{ne} \hat{\beta}^{se})}{N_{ne}} \right] + \left[\sum_{i=1}^{N_{ne}} \frac{F(X_i^{ne} \hat{\beta}^{se})}{N_{ne}} - \sum_{i=1}^{N_{ne}} \frac{F(X_i^{ne} \hat{\beta}^{ne})}{N_{ne}} \right] \quad (2)$$

onde N_j é o tamanho da amostra para a região *j*. Esta expressão alternativa para a decomposição é utilizada porque a média da variável binária observada não necessariamente é a igual a $F(\bar{X}\hat{\beta})$. Na equação (2), o primeiro termo entre parênteses representa a parte do diferencial que é devida a diferenças nas distribuições de *X* entre os grupos, e o segundo termo representa a parte devida a diferenças nos processos de determinação dos níveis de *Y* entre os grupos. O segundo termo também capta a parte da diferença da região devido a diferenças nas dotações imensuráveis ou não-observáveis. De forma semelhante à maioria dos estudos que aplicam a decomposição técnica, não será enfocada a porção da região sem explicação, devido às dificuldades na interpretação de resultados. (FAIRLIE, 2003).

Para calcular a decomposição, define-se \bar{Y}_i como a probabilidade média do resultado binário de interesse para a região *j*, e *F* como a função de distribuição acumulada da distribuição normal padrão. A expressão igualmente válida para a decomposição é:

$$\bar{Y}_{se} - \bar{Y}_{ne} = \left[\sum_{i=1}^{N_{se}} \frac{F(X_i^{se} \hat{\beta}^{se})}{N_{se}} - \sum_{i=1}^{N_{ne}} \frac{F(X_i^{ne} \hat{\beta}^{se})}{N_{ne}} \right] - \left[\sum_{i=1}^{N_{se}} \frac{F(X_i^{se} \hat{\beta}^{se})}{N_{se}} - \sum_{i=1}^{N_{ne}} \frac{F(X_i^{ne} \hat{\beta}^{ne})}{N_{ne}} \right] \quad (3)$$

Neste caso, os coeficientes estimados para o Nordeste ($\hat{\beta}^{ne}$) são utilizados como pesos no primeiro termo da decomposição e a distribuição das variáveis independentes, na região Sudeste (\bar{X}^{se}), como pesos para o segundo termo da equação. Este método alternativo de cálculo da decomposição frequentemente fornece estimativas diferentes, um problema de indexação familiar à metodologia Oaxaca-Blinder. Numa terceira alternativa, usada em Neumark (1988) e Oaxaca e Ransom (1994), o peso é o primeiro termo da decomposição da expressão utilizando estimativas do coeficiente de uma amostra combinada dos dois grupos. Em última análise, a escolha entre estes métodos alternativos de cálculo do primeiro termo da decomposição é difícil e depende da aplicação, com muitos estudos, relatando os resultados de mais de uma especificação. (FAIRLIE, 2003). Desta forma, com vistas a prover maior robustez às análises realizadas neste estudo, as decomposições são realizadas utilizando-se como pesos os resultados estimados para os determinantes da condição de *mismatch* e de adequação das amostras do Nordeste e Sudeste, separadamente, das duas regiões conjuntamente, e para todas as macrorregiões brasileiras.

Os primeiros termos em (2) e (3) fornecem uma estimativa da contribuição das diferenças regionais em todo o conjunto de variáveis independentes para as diferenças na incidência da inadequação no mercado de trabalho entre o Sudeste e o Nordeste. A estimativa da contribuição total é relativamente simples, sendo necessário apenas calcular probabilidades previstas para cada grupo e tomar a diferença entre os valores médios das duas. Identificar a contribuição de diferenças entre os grupos em variáveis específicas para o diferencial regional, no entanto, não é corriqueiro. Por simplicidade, em primeiro lugar assumimos que $N_{ne} = N_{se}$ e que existe um alinhamento *vis-à-vis* entre as observações do Nordeste e Sudeste. Usando estimativas dos coeficientes da equação (1) para uma amostra conjunta das duas regiões ($\hat{\beta}^*$), a contribuição independente de x_i para o diferencial regional pode, então, ser expressa como:

$$\frac{1}{N^{ne}} \sum_{i=0}^{N^{ne}} F(\hat{\alpha}^* + X_{1i}^{se} \hat{\beta}_1^* + X_{2i}^{se} \hat{\beta}_2^*) -$$

$$F(\hat{\alpha}^* + X_{1i}^{ne} \hat{\beta}_1^* + X_{2i}^{se} \hat{\beta}_2^*) \quad (4)$$

De forma similar, a contribuição de x_2 pode ser expressa como:

$$\frac{1}{N^{ne}} \sum_{i=0}^{N^{ne}} F(\hat{\alpha}^* + X_{1i}^{ne} \hat{\beta}_1^* + X_{2i}^{se} \hat{\beta}_2^*) - F(\hat{\alpha}^* + X_{1i}^{ne} \hat{\beta}_1^* + X_{2i}^{ne} \hat{\beta}_2^*) \quad (5)$$

A contribuição de cada variável para a diferença é, portanto, igual à variação na probabilidade média prevista, substituindo-se a distribuição desta variável para a região Nordeste pela distribuição dessa mesma variável na região Sudeste, mantendo as distribuições das demais variáveis constantes. A propriedade útil desta técnica é que a soma das contribuições de variáveis individuais será igual ao total da contribuição de todas as variáveis explicativas constantes na equação (1). O desvio-padrão também pode ser calculado em relação a estas estimativas. Ademais, utiliza-se o método delta para aproximar os desvios-padrão. Para simplificar a notação, reescreve-se (4) como:

$$\hat{D}_1 = \frac{1}{N_{ne}} \sum_{i=1}^{N_{ne}} F(X_i^{se,se} \hat{\beta}^*) - F(X_i^{ne,se} \hat{\beta}^*) \quad (6)$$

A variância de \hat{D}_1 pode ser aproximada como:

$$Var(\hat{D}_1) = \left(\frac{\delta \hat{D}_1}{\delta \hat{\beta}} \right)' Var(\hat{\beta}^*) \left(\frac{\delta \hat{D}_1}{\delta \hat{\beta}^*} \right) \quad (7)$$

onde $\frac{\delta \hat{D}_1}{\delta \hat{\beta}} = \frac{1}{N_{se}} \sum_{i=1}^{N_{se}} f(X_i^{ne,ne} \hat{\beta}) X_i^{ne,ne} - f(X_i^{se,ne} \hat{\beta}) X_i^{se,ne}$ e f é a função de densidade da normal-padrão.

Na prática, as dimensões das amostras dos dois grupos não são as mesmas e é necessário uma disposição de observações *vis-à-vis* entre as amostras para calcular (4), (5) e (7). Neste exemplo, o tamanho da amostra Nordeste é relativamente menor do que o tamanho da amostra Sudeste. Para resolver este problema, primeiro, utilizamos os coeficientes estimados a partir da amostra conjunta

das duas regiões (Nordeste e Sudeste) para calcular a probabilidade predita, \hat{Y}_i , para cada observação na amostra. Em seguida, constrói-se uma subamostra aleatória do Sudeste em tamanho igual à amostra total do Nordeste (N_{ne}) e cada observação nas subamostras do Nordeste e do Sudeste, da amostra total, é então ranqueada separadamente pelas probabilidades preditas e agrupadas segundo seus ranques. Este processo aloca os indivíduos da subamostra construída para região Sudeste colocando-os na parte inferior (superior) de sua distribuição, coincidindo com os indivíduos da amostra da região Nordeste, que têm características observáveis semelhantes na parte inferior (superior) desta distribuição.

A decomposição estimada, obtida a partir deste processo, dependerá da subamostra obtida aleatoriamente para o Sudeste. Virtualmente, os resultados obtidos a partir da decomposição deveriam aproximar-se daquela onde a amostra completa da região Sudeste fosse combinada à amostra da região Nordeste. Um método simples para aproximarmos esta decomposição hipotética é construir um grande número de subamostras aleatórias do Sudeste, correspondendo cada uma destas subamostras aleatórias do Sudeste para amostra do Nordeste e calculando estimativas separadas de cada decomposição. O valor médio das estimativas destas decomposições separadas é calculado e usado para aproximar os resultados para a amostra de todo o Sudeste. Para fins deste trabalho, utilizaram-se 100 subamostras aleatórias da região Sudeste para o cálculo das decomposições a partir das quais foram calculadas as médias das contribuições das variáveis explicativas.

3.3 – Evidências Iniciais: Incidência do Mismatch no Mercado de Trabalho Brasileiro

A aferição da inadequação dos trabalhadores a partir dos níveis constantes na CBO-2002 para cada família ocupacional revela que a incidência dos níveis de inadequação não é homogênea entre as regiões do Brasil, como podemos observar no Gráfico 1. Diferenças no montante de educação média da população de cada região, ou até mesmo mercados de trabalho com dinâmicas diversas devido à presença de setores econômicos particulares em determinadas

áreas, dentre outros fatores, podem ajudar a explicar essa constatação.

De acordo com o Gráfico 1, 49,5% dos trabalhadores brasileiros possuiriam o nível de educação requerida para as suas famílias ocupacionais, revelando uma alta incidência de *mismatch* no mercado de trabalho. Dentro desta categoria, a subeducação atingiria 27,7% dos indivíduos ocupados no mercado de trabalho, enquanto 22,8% deteriam um montante de anos de estudo superior ao máximo requerido pela família ocupacional. O resultado diferente do encontrado por Diaz e Machado (2008) pode ser atribuído à utilização de amostras diferentes e pelo lapso temporal entre elas (IBGE, 2000), bem como pela seleção *ad hoc* das famílias ocupacionais analisadas neste último estudo.

Ao focarmos nos valores para as macrorregiões brasileiras, constatamos que os percentuais aferidos para o *mismatch* nas regiões Sul e Centro-Oeste aproximam-se das médias verificadas para o Brasil. Do mesmo modo, entre estas duas regiões, a incidência de subeducação não apresenta diferença estatisticamente significativa. A região Sudeste apresenta o maior nível de adequação de seus trabalhadores dentre as macrorregiões do país. Outro fator interessante: as regiões Norte e Nordeste apresentam os menores níveis de sobre-educação observados, fator possivelmente relacionado a uma menor média de anos de estudo observada nestas regiões. Contudo, as regiões Nordeste e Sudeste destacam-se por representarem os dois extremos na distribuição do *mismatch*. A região Nordeste apresenta o maior nível verificado de subeducação (32,8%) e a menor proporção de sobre-educação (19,6%), enquanto, na região Sudeste, verificamos o menor índice de subeducação (25%) e o maior percentual de sobre-educados (24%) das macrorregiões brasileiras.

Como mencionamos, diferenças nas características dos trabalhadores de cada região (capital humano, demográficas etc.) e na distribuição de setores de atividade ou no grau de formalização da economia poderiam ajudar a explicar o diferencial observado entre o Nordeste e o Sudeste. De fato, não se dispõe na literatura de evidências acerca da contribuição das estruturas de oferta e da demanda por trabalho para o

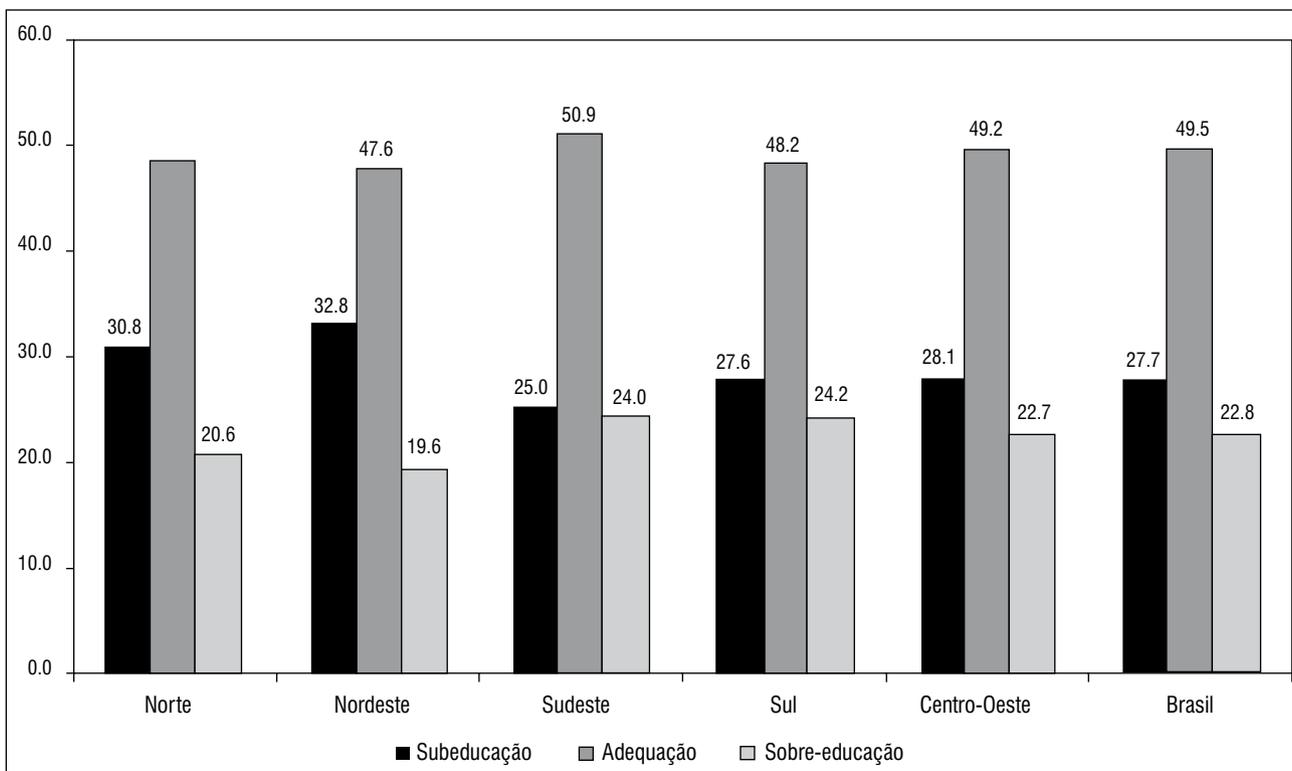


Gráfico 1 – Percentual de Subeducados, Adequados e Sobre-educados para as Macrorregiões do Brasil.

Fonte: Elaboração Própria dos Autores a partir dos Dados da PNAD de 2007.

maior nível de adequação verificado na região Sudeste (50,9%) e o menor na região Nordeste (47,6%). Assim, mercados de trabalho com maior grau de formalização ou ramos de atividade mais dinâmicos em uma das regiões poderiam explicar, em parte, um nível de *mismatch* observado até 7% superior na região Nordeste, em relação ao Sudeste. Desta forma, nas seções seguintes, procura-se determinar quais características individuais e ocupacionais estão associadas à probabilidade de *mismatch* dos indivíduos no mercado de trabalho e qual a importância destas para os diferenciais observados entre as regiões Nordeste e Sudeste do Brasil.

4 – DETERMINANTES DO MISMATCH NO MERCADO DE TRABALHO: EVIDÊNCIAS PARA O BRASIL

Como vimos, apesar de os trabalhos pioneiros de Santos (2002) e Diaz e Machado (2008) apresentarem evidências acerca dos impactos da inadequação

do nível de escolaridade sobre os rendimentos dos trabalhadores brasileiros, não existem evidências na literatura brasileira sobre quais características observáveis poderiam elevar a probabilidade de os trabalhadores estarem sob tal condição. Deste modo, como primeiro passo em nossa investigação, buscamos determinar a importância das características individuais e ocupacionais sobre a probabilidade de *mismatch* entre o nível de educação formal requerido para cada família ocupacional e aquele detido pelo trabalhador.

Assim, a Tabela 3 apresenta os efeitos marginais estimados a partir da equação (1) para a probabilidade de *mismatch* (indivíduos representativos), onde a variável dependente assume valor igual a 1 (um) se o montante de anos de estudo do indivíduo está entre os valores mínimo e máximo determinados para sua família ocupacional (indivíduo com escolaridade adequada) e igual a 0 (zero), caso contrário (*mismatch*). Os determinantes são estimados, separadamente, para os indivíduos ocupados na

região Nordeste e Sudeste, conjuntamente para as duas regiões e para todas as regiões brasileiras. Os indivíduos representativos de cada amostra possuem características observáveis muito próximas, diferindo apenas no tempo que estão no mercado de trabalho⁴ (diferença pouco superior a um ano entre as médias). Assim, por simplificação, doravante trataremos o indivíduo representativo de todos os modelos estimados como homem, com ensino médio completo, casado e sem filhos, não-migrante, com 35 anos de idade, média de 20 anos de experiência no mercado de trabalho, que se declarou de cor/raça não-branca e que residia em área urbana e não metropolitana.

Ademais, como nosso objetivo principal está em aferir as contribuições individuais das características observáveis sobre os diferenciais de *mismatch* observados entre as regiões, os pesos estimados (obtidos dos modelos *probit*) para as variáveis podem influir sobre os resultados obtidos no processo de decomposição não-linear. Assim, seguindo a metodologia proposta por Fairlie (2003), estimaremos os determinantes de *mismatch* (bem como de sub e sobre-educação) a partir das quatro amostras acima apontadas: Nordeste, Sudeste, as duas regiões combinadas e Brasil. Este esforço permite-nos tanto aferir quais os determinantes do *mismatch* entre as regiões e suas prováveis diferenciações como prover maior robustez às decomposições conduzidas adiante, eliminando possíveis vieses advindos da escolha dos pesos inseridos nas equações (4), (5) e (6).

Inicialmente, os resultados estimados para todos os modelos sugerem que a probabilidade de *mismatch* é reduzida para os indivíduos que concluíram alguma fase da educação formal, em comparação aos indivíduos com nenhuma educação formal ou ensino fundamental incompleto (variável omitida). A probabilidade de adequação do nível de educação ao exigido para cada família ocupacional cresceria de 31,2 (Sudeste) até 35,7% (Nordeste) para indivíduos com ensino fundamental completo. Os sinais negativos para os

efeitos marginais estimados para as variáveis ensino médio incompleto (até -11,5%) e ensino superior incompleto (até -34,7%) sugerem que a probabilidade de adequação seria decrescente para os indivíduos com educação formal mais elevada, mas que não concluíram com êxito a última etapa de formação que cursaram. Assim, indivíduos que não concluiriam estas duas etapas da educação formal teriam uma probabilidade maior de *mismatch* do que aqueles que não concluíram o ensino fundamental ou detinham menos de um ano de estudo. Desta forma, as evidências colhidas sugerem que quanto maior o nível de educação do trabalhador, maior a probabilidade de que este se encontre em uma família ocupacional adequada ao seu montante de anos de estudos, especialmente para aqueles que concluíram com sucesso o curso mais elevado que frequentaram.

A experiência do indivíduo no mercado de trabalho tem efeito significativo sobre a probabilidade de *mismatch* dos indivíduos. Assim, indivíduos com maior tempo no mercado teriam uma probabilidade esperada menor de possuírem um nível de educação adequado àquele demandado pela família ocupacional. A condição de cor/raça apresentou resultado significativo, elevando a probabilidade de um *match* mais adequado (exceto para a região Nordeste). A condição de migrante, o estado civil do indivíduo e a participação em sindicatos parecem não afetar a probabilidade de *mismatch* do nível de educação no mercado de trabalho. Por outro lado, as mulheres teriam uma probabilidade esperada maior de se encontrar em ocupações adequadas ao seu montante de anos de estudo, quando comparadas aos homens. Este efeito seria verificado tanto para as duas regiões analisadas como para a amostra composta para todas as macrorregiões, sendo mais relevante na região Nordeste onde este efeito seria de até 18,8% sobre a probabilidade dos homens.

Conforme esperado *a priori*, as regiões metropolitanas incrementariam a probabilidade de o indivíduo encontrar uma ocupação adequada ao seu nível de educação formal. Particularmente na região Nordeste, residir em uma região metropolitana teria um efeito marginal de até 3,5% sobre a probabilidade de o indivíduo ter um nível de educação adequado à ocupação. Por hipótese, este resultado estaria relacionado a mercados de trabalho mais dinâmicos

⁴ O indivíduo representativo da região Sudeste possui a maior média de tempo no mercado de trabalho com 21,6 anos, enquanto o indivíduo representativo da região Nordeste apresenta uma média igual a 20,3 anos. Os indivíduos representativos da amostra conjunta destas duas regiões e da amostra para o Brasil possuem uma média igual a 21,1 anos de experiência e residem na região Sudeste.

Tabela 3 – Efeitos marginais^a dos determinantes da adequação para as regiões Nordeste, Sudeste e Brasil em 2007 (indivíduo representativo) – Modelo *Probit*

	Mod. 1	Mod. 2	Mod. 3	Mod. 4
	NE	SE	NE/SE	BR
Ensino fundamental completo	***0.3569 (0.0085)	***0.3119 (0.0076)	***0.3341 (0.0056)	***0.3336 (0.0043)
Ensino médio incompleto	***-0.0615 (0.0126)	***-0.1505 (0.0121)	***-0.1052 (0.0087)	***-0.1146 (0.0066)
Ensino médio completo	***0.2792 (0.0076)	***0.2114 (0.0073)	***0.2452 (0.0052)	***0.2423 (0.0041)
Ensino superior incompleto	***-0.2522 (0.0129)	***-0.3990 (0.0089)	***-0.3337 (0.0076)	***-0.3475 (0.0055)
Ensino superior completo	***0.1372 (0.0123)	***0.0582 (0.01)	***0.0918 (0.0077)	***0.0969 (0.006)
Experiência	-0.0004 (0.0009)	***-0.0022 (0.0008)	** -0.0015 (0.0006)	***-0.0021 (0.0005)
Experiência ao quadrado	***-0.0058 (0.0018)	-0.0005 (0.0016)	** -0.0024 (0.0012)	*-0.0018 (0.0009)
Branco	0.0097 (0.0068)	***0.0241 (0.0058)	***0.0180 (0.0044)	***0.0218 (0.0035)
Migrante	0.0249 (0.0167)	0.0148 (0.0163)	*0.0199 (0.0117)	0.0101 (0.0081)
Mulher	***0.1878 (0.0068)	***0.1609 (0.0061)	***0.1728 (0.0045)	***0.1659 (0.0035)
Casamento	*-0.0132 (0.0072)	-0.0002 (0.0066)	-0.0042 (0.0048)	-0.0038 (0.0037)
Região metropolitana	***0.0349 (0.0065)	***0.0234 (0.0057)	***0.0275 (0.0043)	***0.0271 (0.0033)
Área rural	***-0.0579 (0.0116)	0.0163 (0.0138)	***-0.0316 (0.0088)	***-0.0292 (0.0065)
Sindicalizado	0.0097 (0.0089)	-0.0009 (0.0078)	0.0033 (0.0059)	0.0068 (0.0045)
Presença de filhos	Sim	Sim	Sim	Sim
Tipo de ocupação	Sim	Sim	Sim	Sim
Ramo de atividade	Sim	Sim	Sim	Sim
Macrorregião			Sim	Sim
Pred. ($y=1$)	0.4850	0.5072	0.4981	0.4941
Log <i>likelihood</i>	-17807.3	-22212.0	-40187.6	-67185.3
Pseudo R2	0.1735	0.1494	0.157	0.1585
Prob > Chi2	0.0000	0.0000	0.0000	0.0000
Número de observações	31092	37687	68779	115182

Fonte: Elaboração Própria dos Autores a partir dos Microdados da PNAD de 2007.

Notas: * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$. Erros-padrão estimados entre parênteses. Regressões ajustadas para presença de filhos (menores de 06 anos e entre 07 e 14 anos), tipos de ocupação (carteira assinada e servidor público), ramos de atividade (Setor Agrícola e Extrativo; Indústrias de Transformação; Construção; Comércio e Serviços Pessoais; Serviços Financeiros etc.; Administração Pública e Serviços Sociais; Outros Serviços) e macrorregiões brasileiras. ^(a) dy/dx para um mudança discreta (de 0 para 1) na variável *dummy*, exceto para as variáveis experiência e experiência ao quadrado que apresentam distribuição contínua

e com uma maior disponibilidade de ocupações, além de custos de deslocamento menores e maior difusão de conhecimento e informação. Nesta perspectiva, o resultado significativo estatisticamente e negativo (exceto para a região Sudeste) para as áreas rurais suporta a hipótese levantada, revelando uma penalidade sobre a probabilidade de adequação, quando comparado aos indivíduos residentes em zonas urbanas.

Entretanto, indivíduos que não possuam o nível de educação requerido para a família de ocupação podem ser categorizados como subeducados ou sobre-educados, com consequências diferentes sobre seus rendimentos. (HARTOG, 2000). Assim, uma vez observadas quais características individuais, poderiam estar relacionadas à probabilidade de o indivíduo ter um *match* adequado para seu montante de anos de estudo, passamos a investigar quais os determinantes da subeducação e da sobre-educação para o mercado de trabalho brasileiro. A Tabela 4 apresenta os resultados para os efeitos marginais da condição de subeducação para o indivíduo representativo descrito acima. Estimados a partir da equação (1), onde a variável explicada assume valor igual a 1 (um) se o indivíduo possui um montante de anos de estudo menor que o requerido para a família ocupacional e igual a 0 (zero), caso contrário.

Os efeitos do nível de educação sobre a probabilidade de o indivíduo ser classificado como subeducado apresenta o resultado esperado e é coerente com os resultados encontrados na Tabela 3. O montante de anos de estudo estaria negativamente relacionado à probabilidade de subeducação em todas as amostras. Novamente, este efeito seria mais relevante para aqueles indivíduos que concluíram com sucesso o último curso que frequentaram. Por exemplo, para todas as regiões brasileiras (Mod. 4), indivíduos com ensino médio incompleto teriam um efeito marginal de -17,9% de serem subeducados, enquanto, para aqueles que completaram o ensino médio, esta probabilidade seria de -41,1%, comparados aos indivíduos que não terminaram o ensino fundamental ou não possuíam educação formal.

Quanto ao tempo de permanência no mercado de trabalho, os resultados estimados parecem sugerir uma relação linear entre esta e a probabilidade de ser

categorizado como subeducado. Onde, quão maior o período no mercado de trabalho, mais elevadas seriam as chances de se encontrar numa família ocupacional cuja educação requerida é maior que aquela detida pelo indivíduo. A cor/raça declarada parece não influenciar a probabilidade de subeducação, enquanto a condição de migrante, apenas na região Sudeste, teria um efeito marginal de -2,5% sobre o risco de o trabalhador ser categorizado como subeducado. As mulheres, em comparação aos homens, teriam uma probabilidade até -10,7% de serem subeducadas, o que pode estar relacionado, dentre outros fatores, a maior média de educação verificada entre as mulheres.

O casamento parece ter um efeito relevante sobre a probabilidade de subeducação, particularmente na região Nordeste. Nesta região, indivíduos casados teriam uma probabilidade até 1,5% maior de serem subeducados no mercado de trabalho. As regiões metropolitanas teriam um efeito diferente sobre a probabilidade de subeducação para as regiões Nordeste (-2,7%) e Sudeste (0,9%), enquanto, para o Brasil, este efeito seria negativo, reduzindo a probabilidade de *mismatch*, ao contrário das áreas rurais, que elevariam a probabilidade de subeducação na região Nordeste (3,8%) e no Brasil (1,7%), para região Sudeste, não haveria efeito estatisticamente relevante. Por fim, trabalhadores vinculados a algum sindicato laboral teriam uma probabilidade esperada de subeducação maior que a daqueles não-filiados a sindicatos, com um efeito marginal de até 3,0% (região Sudeste).

Do mesmo modo que investigamos os determinantes da subeducação, a Tabela 5 apresenta os efeitos marginais calculados para a condição de sobre-educado, onde a variável dependente toma valor igual a 1 (um), caso o trabalhador detenha mais educação formal que o limite máximo requerido para a família ocupacional, sendo igual a 0 (zero) no caso contrário. Em conformidade com os diversos estudos nesta área, o nível de educação formal também tem um efeito positivo sobre a probabilidade de o indivíduo ser tipificado como sobre-educado no mercado de trabalho. (MCGOLDRICK; ROBST, 1996; LINSLEY, 2005). Ademais, este risco parece mais relevante para os indivíduos que não concluíram o último curso que frequentaram. Para o Brasil (Mod. 4), os indivíduos

Tabela 4 – Efeitos Marginais^a dos Determinantes da Subeducação para as Regiões Nordeste, Sudeste e Brasil em 2007 (Indivíduo Representativo) – Modelo *Probit*

	Mod. 1	Mod. 2	Mod. 3	Mod. 4
	NE	SE	NE/SE	BR
Ensino fundamental completo	***-0.2223 (0.0036)	***-0.1845 (0.003)	***-0.2021 (0.0023)	***-0.2053 (0.0018)
Ensino médio incompleto	***-0.2047 (0.0038)	***-0.1546 (0.0034)	***-0.1779 (0.0025)	***-0.1798 (0.0019)
Ensino médio completo	***-0.4540 (0.0044)	***-0.3892 (0.0039)	***-0.4185 (0.0029)	***-0.4110 (0.0022)
Ensino superior incompleto	***-0.2134 (0.0036)	***-0.1751 (0.003)	***-0.1923 (0.0023)	***-0.1966 (0.0018)
Ensino superior completo	***-0.2590 (0.0037)	***-0.2450 (0.0031)	***-0.2521 (0.0024)	***-0.2570 (0.0018)
Experiência	0.0008 (0.0008)	***-0.0020 (0.0007)	-0.0005 (0.0005)	*-0.0007 (0.0004)
Experiência ao quadrado	***0.0052 (0.0015)	***0.0064 (0.0012)	***0.0054 (0.001)	***0.0056 (0.0007)
Branco	0.0067 (0.0061)	0.0072 (0.0044)	*0.0067 (0.0036)	*0.0048 (0.0029)
Migrante	0.0050 (0.0148)	** -0.0248 (0.012)	-0.0113 (0.0094)	-0.0072 (0.0067)
Mulher	***-0.1075 (0.006)	***-0.0786 (0.0049)	***-0.0913 (0.0038)	***-0.0881 (0.003)
Casamento	***0.0154 (0.0061)	*0.0086 (0.0051)	**0.0087 (0.0039)	**0.0059 (0.003)
Região metropolitana	***-0.0273 (0.0057)	**0.0093 (0.0044)	*-0.0059 (0.0035)	***-0.0075 (0.0028)
Área rural	***0.0388 (0.0101)	-0.0085 (0.0095)	***0.0231 (0.007)	***0.0168 (0.0051)
Sindicalizado	**0.0190 (0.0084)	***0.0303 (0.0068)	***0.0266 (0.0053)	***0.0241 (0.0041)
Presença de filhos	Sim	Sim	Sim	Sim
Tipo de ocupação	Sim	Sim	Sim	Sim
Ramo de atividade	Sim	Sim	Sim	Sim
Macrorregião			Sim	Sim
Pred.(y=1)	0.2032	0.1714	0.1863	0.1883
Log likelihood	-11422.1	-14205.7	-25802.0	-43466.6
Pseudo R2	0.4050	0.3350	0.3658	0.3620
Prob>Chi2	0.0000	0.0000	0.0000	0.0000
Número de observações	31092	37687	68779	115182

Fonte: Elaboração Própria dos Autores a Partir dos Microdados da PNAD de 2007.

Notas: * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$. Erros-padrão estimados entre parênteses. Regressões ajustadas para presença de filhos (menores de 06 anos e entre 07 e 14 anos), tipos de ocupação (carteira assinada e servidor público), ramos de atividade (Setor Agrícola e Extrativo; Indústrias de Transformação; Construção; Comércio e Serviços Pessoais; Serviços Financeiros etc.; Administração Pública e Serviços Sociais; Outros Serviços) e macrorregiões brasileiras. ^(a) dy/dx para um mudança discreta (de 0 para 1) na variável *dummy*, exceto para as variáveis experiência e experiência ao quadrado que apresentam distribuição contínua.

Tabela 5 – Efeitos Marginais dos Determinantes da Sobre-educação para as Regiões Nordeste, Sudeste e Brasil em 2007 (Indivíduo Representativo) – Modelo Probit

	Mod. 1	Mod. 2	Mod. 3	Mod. 4
	NE	SE	NE/SE	BR
Ensino fundamental completo	***0.3003 (0.0158)	***0.3045 (0.0155)	***0.2974 (0.011)	***0.3142 (0.0084)
Ensino médio incompleto	***0.7474 (0.0101)	***0.7746 (0.009)	***0.7624 (0.0067)	***0.7693 (0.005)
Ensino médio completo	***0.5331 (0.0089)	***0.6067 (0.009)	***0.5708 (0.0063)	***0.5825 (0.0049)
Ensino superior incompleto	***0.8742 (0.0053)	***0.8858 (0.0038)	***0.8812 (0.0031)	***0.8855 (0.0022)
Ensino superior completo	***0.7558 (0.0104)	***0.7987 (0.0082)	***0.7785 (0.0064)	***0.7776 (0.0049)
Experiência	***0.0016 (0.0005)	***0.0039 (0.0006)	***0.0027 (0.0004)	***0.0031 (0.0003)
Experiência ao quadrado	***-0.0043 (0.0012)	***-0.0067 (0.0012)	***-0.0054 (0.0009)	***-0.0060 (0.0007)
Branco	** -0.0084 (0.0038)	***-0.0255 (0.0039)	***-0.0171 (0.0027)	***-0.0193 (0.0022)
Migrante	** -0.0182 (0.0084)	0.0027 (0.0105)	-0.0085 (0.0068)	-0.0040 (0.0049)
Mulher	***-0.0673 (0.0039)	***-0.0650 (0.0039)	***-0.0661 (0.0028)	***-0.0634 (0.0022)
Casamento	-0.0029 (0.0043)	*-0.0085 (0.0045)	-0.0050 (0.0031)	-0.0020 (0.0024)
Região metropolitana	-0.0057 (0.0038)	***-0.0247 (0.0037)	***-0.0155 (0.0026)	***-0.0134 (0.0021)
Área rural	0.0112 (0.0078)	0.0102 (0.0106)	0.0101 (0.0064)	***0.0145 (0.0048)
Sindicalizado	** -0.0110 (0.0046)	***-0.0163 (0.0045)	***-0.0146 (0.0032)	***-0.0161 (0.0025)
Presença de filhos	Sim	Sim	Sim	Sim
Tipo de ocupação	Sim	Sim	Sim	Sim
Ramo de atividade	Sim	Sim	Sim	Sim
Macrorregião			Sim	Sim
Pred.(y=1)	0.1080	0.1340	0.1225	0.1251
Log likelihood	-11225.1	-14775.7	-26081.2	-43768.6
Pseudo R2	0.2816	0.2802	0.2797	0.2835
Prob>Chi2	0.0000	0.0000	0.0000	0.0000
Número de observações	31092	37687	68779	115182

Fonte: Elaboração Própria dos Autores a partir dos Microdados da PNAD de 2007.

Notas: * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$. Erros-padrão estimados entre parênteses. Regressões ajustadas para presença de filhos (menores de 06 anos e entre 07 e 14 anos), tipos de ocupação (carteira assinada e servidor público), ramos de atividade (Setor Agrícola e Extrativo; Indústrias de Transformação; Construção; Comércio e Serviços Pessoais; Serviços Financeiros etc.; Administração Pública e Serviços Sociais; Outros Serviços) e macrorregiões brasileiras. (*) dy/dx para uma mudança discreta (de 0 para 1) na variável *dummy*, exceto para as variáveis experiência e experiência ao quadrado, que apresentam distribuição contínua.

que não concluíram o ensino superior teriam um efeito marginal de 88,5% sobre a probabilidade de sobre-educação, ao passo que, entre aqueles que completaram o ensino superior, esta percentagem cai para 77,8%, em relação ao grupo de referência. Dado que indivíduos mais habilidosos teriam menores custos para acumular anos adicionais de educação formal que indivíduos menos habilidosos, de acordo com Cahuc e Zylberberg (2004), este efeito pode sugerir também que indivíduos mais habilidosos teriam maior probabilidade de serem sobre-educados.

Os coeficientes estimados para experiência e experiência ao quadrado, estatisticamente significantes a 1%, incluídos para captar uma relação não-linear, parecem indicar um relação direta entre experiência e inadequação onde quanto maior o número de anos no mercado de trabalho, maior o risco de o indivíduo ser sobre-educado. Este resultado sugere que, para as regiões do Brasil, a condição de sobre-educado não parece ser uma estratégia de inserção mais cedo no mercado de trabalho, isto é, o passar do tempo no mercado de trabalho não parece fazer o indivíduo alcançar postos de trabalho mais apropriados à sua qualificação.

As pessoas que se declararam de cor/raça branca também teriam uma chance menor de estar na condição de sobre-educados do que pessoas que se declaram de outra cor/raça, com efeitos marginais negativos para as regiões Nordeste (-0,8%) e Sudeste (-2,5%). Da mesma forma, os trabalhadores sindicalizados teriam uma probabilidade menor de serem categorizados como sobre-educados. A condição de migração apresenta efeito marginal significativo apenas na região Nordeste, onde o migrante teria uma probabilidade até 1,8% menor de ser sobre-educado do que os trabalhadores não-migrantes da região. Os efeitos marginais para o estado civil e áreas rurais não apresentam resultados robustos acerca do seu impacto sobre a probabilidade de *mismatch*.

As mulheres teriam, outra vez, um risco menor de serem classificadas como sobre-educadas, em comparação aos homens, em todas as regiões brasileiras. Mulheres teriam um efeito marginal médio de -6,5% sobre a probabilidade de *mismatch*. Por fim,

verificamos que as regiões metropolitanas reduzem o risco de sobre-educação dos indivíduos, o que corrobora hipótese de melhores *matches* em mercados mais amplos e dinâmicos. Trabalhadores que residiam em áreas urbanas teriam um efeito marginal de até -2,5% para a região Sudeste e de até -1,3% para todas as macrorregiões brasileiras. Contudo, tomando-se apenas os indivíduos ocupados da região Nordeste não se verifica um efeito estatisticamente significativo para as regiões metropolitanas.

5 – DETERMINANTES DOS DIFERENCIAIS REGIONAIS DE *MISMATCH* NO BRASIL

Na seção anterior, verificamos quais características observáveis dos indivíduos influiriam na probabilidade de *mismatch* do nível de educação destes no mercado de trabalho, tanto para as regiões Nordeste e Sudeste como para o Brasil. Foi visto que a região Nordeste apresenta maior grau de *mismatch* ou inadequação do trabalho em seu mercado regional. Nesta seção, com base nos resultados estimados para a equação (1), seguindo a metodologia de Fairlie (1999, 2003), decomposmos a importância de cada característica observável sobre o diferencial de *mismatch* entre as regiões Nordeste e Sudeste. O objetivo é verificar em que medida as variáveis de oferta (características dos indivíduos) e demanda (tipo de ocupação e ramos de atividades) explicam as disparidades de ajustamento entre tais regiões.

Na busca de maior robustez, a decomposição não-linear é realizada tomando-se como base cada um dos modelos *probit* estimados. Como ressalta Fairlie (2003), a utilização dos diversos pesos estimados para a determinação do *mismatch* nas decomposições afastaria possíveis vieses resultantes da escolha do vetor de coeficientes, trazendo mais robustez aos resultados obtidos. Ademais, na decomposição, em cada replicação da amostra, a ordem das variáveis foi aleatoriamente determinada como forma de mitigar o efeito da ordem das variáveis sobre os resultados obtidos.

Assim, na Tabela 6, são apresentados os resultados da decomposição não-linear para o diferencial de adequação entre as regiões Nordeste e Sudeste. A

contribuição de grupos de variáveis, especialmente *dummies* (presença de filhos, tipo de ocupação e ramo de atividade) e experiência (experiência e experiência ao quadrado), é calculada pela substituição conjunta da distribuição de todas as variáveis de cada categoria. Os resultados obtidos são similares entre as especificações modeladas; no entanto, por parcimônia,

concentraremos nossa análise naquelas variáveis que se mostraram significantes no nível de confiança de 5%, em ao menos duas decomposições.

Pela Tabela 6, a proporção de trabalhadores adequadamente alocados na região Sudeste é dois pontos percentuais superior àquela observada na região

Tabela 6 – Contribuição Percentual das Características Observáveis dos Indivíduos para o Diferencial de Adequação entre as Regiões Sudeste e Nordeste

	Mod. 1		Mod. 2		Mod. 3		Mod. 4	
Amostra usada no modelo probit:	NE		SE		NE/SE		BR	
Parcela de adequação observada:								
Sudeste	0.5113		0.5113		0.5113		0.5113	
Nordeste	0.4905		0.4905		0.4905		0.4905	
Diferencial predito entre as regiões	0.0208		0.0208		0.0208		0.0208	
Parcela explicada pelas variáveis	0.0078		0.0060		0.0086		0.0082	
Contribuição para o diferencial devido a:								
Ensino fundamental completo	41.57	**	36.26	**	42.89	**	34.56	**
Ensino médio incompleto	3.10	**	7.39	**	4.90	**	5.36	**
Ensino médio completo	18.84	**	16.78	**	15.04	**	21.37	**
Ensino superior incompleto	-5.64	**	-15.53	**	-4.62	**	-6.85	**
Ensino superior completo	25.45	**	11.30	**	17.69	**	18.06	**
Experiência	-15.67	**	-12.40	**	-13.59	**	-14.87	**
Branco	9.47		24.84	**	18.04	**	21.92	**
Migrante	-0.53		-0.33		-0.44		-0.22	
Mulher	-4.11	**	-3.41	**	-4.81	**	-3.62	**
Casamento	-0.45		-0.01		-0.17		-0.14	
Presença de filhos	-0.93		3.40	**	1.46		0.67	
Região metropolitana	-0.01		0.22	**	0.10		0.18	**
Área Rural	9.61	**	-3.18		5.71	**	5.36	**
Sindicalizado	0.04		-0.01		0.02		0.05	
Tipo de ocupação	2.22		-6.43		-0.32		-4.15	
Ramo de Atividade	-45.72	**	-29.85	**	-40.78	**	-37.81	**
Total explicado (%)	37.61		28.93		41.16		39.70	

Fonte: Elaboração Própria dos Autores a partir dos Microdados da PNAD de 2007.

Notas: ** $p \leq 0,05$. As contribuições estimadas reportadas são os valores médios obtidos para as decomposições usando 100 subamostras da região Sudeste. Regressões ajustadas para experiência (experiência e experiência ao quadrado), presença de filhos (menores de 06 anos e entre 07 e 14 anos), tipos de ocupação (carteira assinada e servidor público), ramos de atividade (Setor Agrícola e Extrativo; Indústrias de Transformação; Construção; Comércio e Serviços Pessoais; Serviços Financeiros etc.; Administração Pública e Serviços Sociais; Outros Serviços) e macrorregiões brasileiras (não computadas na decomposição acima).

Tabela 7 – Contribuição Percentual das Características Observáveis dos Indivíduos para o Diferencial de Subeducação entre as Regiões Nordeste e Sudeste

	Mod. 1		Mod. 2		Mod. 3		Mod. 4	
Amostra usada no modelo probit:	NE		SE		NE/SE		BR	
Parcela de subeducação observada:								
Sudeste	0.2541		0.2541		0.2541		0.2541	
Nordeste	0.3079		0.3079		0.3079		0.3079	
Diferencial predito entre as regiões	-0.0537		-0.0537		-0.0537		-0.0537	
Parcela explicada pelas variáveis	-0.0360		-0.0293		-0.0335		-0.0332	
Contribuição para o diferencial devido a:								
Ensino fundamental completo	27.50	**	24.98	**	29.22	**	29.30	**
Ensino médio incompleto	3.34	**	-0.85		-0.25		2.29	**
Ensino médio completo	7.36	**	14.71	**	10.68	**	14.12	**
Ensino superior incompleto	10.84	**	10.27	**	7.26	**	11.70	**
Ensino superior completo	39.95	**	28.83	**	34.54	**	25.45	**
Experiência	-6.73	**	-2.77	**	-4.66	**	-4.37	**
Branco	-2.18		-2.90		-2.42		-1.78	
Migrante	0.03		-0.19		-0.08		-0.05	
Mulher	3.13	**	1.81	**	2.28	**	1.14	**
Casamento	-0.34		-0.20		-0.25		-0.15	
Presença de filhos	-0.28		0.09		0.04		-0.03	
Região metropolitana	0.54	**	-0.23		0.14		0.19	**
Área rural	2.98	**	-0.84		2.06	**	1.50	**
Sindicalizado	-0.06		-0.26		0.03		-0.09	
Tipo de ocupação	-8.19	**	-9.80	**	-7.41	**	-7.05	**
Ramo de atividade	-10.89	**	-8.13	**	-8.64	**	-10.40	**
Total explicado (%)	66.94		54.52		62.24		61.77	

Fonte: Elaboração Própria dos Autores a partir dos Microdados da PNAD 2007.

Notas: ** $p \leq 0,05$. As contribuições estimadas reportadas são os valores médios obtidos para as decomposições usando 100 subamostras da região Sudeste. Regressões ajustadas para experiência (experiência e experiência ao quadrado), presença de filhos (menores de 06 anos e entre 07 e 14 anos), tipos de ocupação (carteira assinada e servidor público), ramos de atividade (Setor Agrícola e Extrativo; Indústrias de Transformação; Construção; Comércio e Serviços Pessoais; Serviços Financeiros etc.; Administração Pública e Serviços Sociais; Outros Serviços) e macrorregiões brasileiras (não-computadas na decomposição acima).

Nordeste (diferença estatisticamente significativa), representando um nível de *mismatch* até 4,2% superior no Nordeste. Deste diferencial regional, as características observáveis dos indivíduos destas duas regiões explicam, em média, 36,8% deste diferencial. Nesta contribuição, as condições de oferta de trabalho seriam relevantes na explicação do diferencial observado. Como esperado, o nível de educação dos indivíduos é um fator que explica grande parte do

diferencial de adequação entre as regiões. Uma maior proporção de indivíduos que concluíram o último curso que frequentaram na região Sudeste, em comparação à região Nordeste, seria um dos fatores relevantes para a explicação do diferencial. Por exemplo, a proporção de indivíduos com ensino fundamental completo explicaria de 34,5 a 42,9% do diferencial observado. A diferença na proporção de indivíduos com ensino superior incompleto entre o Nordeste (4,7%) e o Sudeste (6,0%)

teria um efeito negativo sobre o diferencial que giraria entre -4,6 e -15,5%.

A proporção de indivíduos que se declararam de cor/raça branca influiria de 18% a 24,8% sobre o diferencial de adequação observado, o que é explicado pela proporção muito superior destes indivíduos na região Sudeste (56%), quando comparada com a região Nordeste (30,3%), e uma probabilidade maior de estes estarem alocados adequadamente no mercado de trabalho. No quesito gênero, a distribuição das mulheres entre as duas regiões mitigaria o diferencial entre -3,4 e -4,8% a favor da região Nordeste. Do mesmo modo, uma concentração maior da população em áreas rurais na região Nordeste (9,4%) do que no Sudeste (4,6%), na amostra, explicaria entre 5,3 a 9,6% do diferencial observado. A migração, o estado civil e a distribuição dos indivíduos nas áreas metropolitanas, em tese, não auxiliariam na explicação do diferencial de adequação entre o Nordeste e o Sudeste.

As condições de demanda de trabalho também apresentariam um fator relevante na explicação do diferencial de *mismatch* observado entre as regiões. Dentro dos fatores de demanda, a distribuição dos indivíduos ocupados entre os ramos de atividade nas duas macrorregiões parece relevante para a explicação do diferencial regional. Entre -29,8 e -45,7% do diferencial de adequação entre as regiões poderiam ser creditados à distribuição dos indivíduos nos ramos de atividade da economia. Basicamente, pelos efeitos de uma concentração maior da mão-de-obra ocupada na região Nordeste em ramos de atividade com efeitos marginais positivos sobre a probabilidade de adequação ou em ramos com efeitos negativos menores (em módulo). Por fim, diferenças entre a proporção de sindicalizados e na distribuição entre os tipos de ocupação não teriam significância estatística na explicação do diferencial de adequação.

Para o caso específico de *mismatch* associado à situação de subeducação (*undereducation*), a Tabela 7 apresenta a decomposição não-linear do diferencial entre as regiões Nordeste e Sudeste. As contribuições reportadas são calculadas com base nos modelos *probit* para condição de subeducados, estimados na Tabela 4 para o Nordeste, Sudeste, Nordeste e Sudeste conjuntamente e todas as macrorregiões

brasileiras. A incidência observada da subeducação seria maior na região Nordeste (30,8%) do que na região Sudeste (25,4%), reportando um diferencial de 5,4% de subeducação. Assim, a região Nordeste teria uma incidência de subeducação até 21,2% maior que a observada no Sudeste. As decomposições indicam que as diferenças nas características observáveis dos indivíduos entre as macrorregiões explicariam entre 54,5 e 66,9% deste diferencial observado.

Assim, como no caso da decomposição do diferencial de adequação, as condições de oferta de trabalho seriam relevantes na explicação deste diferencial. O nível de escolaridade dos indivíduos teria uma importância relevante na explicação do diferencial de subeducação entre as regiões. Neste aspecto, as diferenças regionais na distribuição de indivíduos com ensino fundamental completo e ensino superior completo entre as regiões seriam as mais relevantes para compreensão do diferencial observado, explicando até 29,3 e 39,9% deste, respectivamente. A experiência no mercado de trabalho (maior na região Sudeste) também ajudaria a explicar parte deste diferencial. Caso os indivíduos da região Sudeste possuíssem a mesma distribuição das variáveis de experiência, o diferencial seria reduzido entre -2,8 e -6,7%. Enquanto a distribuição dos indivíduos por gênero e entre as zonas rurais e urbanas teria um efeito positivo sobre o diferencial.

Por fim, se as distribuições dos ocupados entre os tipos de ocupação e entre os ramos de atividade da região Nordeste fossem replicadas para o Sudeste, o diferencial observado de subeducação entre as regiões seria mitigado, respectivamente, em até -9,8% e -10,8%. Ou seja, as diferentes estruturais regionais de demanda por trabalho são responsáveis por até cerca de 20% do diferencial regional de subeducação observado.

Na Tabela 8, apresentam-se os resultados de decomposições das diferenças regionais entre as duas regiões, agora, para o caso de *mismatch* associado à condição de sobre-educação. Relembre-se que a parcela de sobre-educados observada é de cerca de 23,4% da mão-de-obra ocupada na região Sudeste e de 20,1%, na região Nordeste. O diferencial de 3,3 pontos percentuais representa, pois, uma

Tabela 8 – Contribuição Percentual das Características Observáveis dos Indivíduos para o Diferencial de Sobre-educação entre as Regiões Sudeste e Nordeste

	Mod. 1	Mod. 2	Mod. 3	Mod. 4
Amostra usada no modelo probit:	NE	SE	NE/SE	BR
Parcela de sobre-educação observada:				
Sudeste	0.2345	0.2345	0.2345	0.2345
Nordeste	0.2016	0.2016	0.2016	0.2016
Diferencial predito entre as regiões	0.0330	0.0330	0.0330	0.0330
Parcela explicada pelas variáveis	0.0276	0.0233	0.0253	0.0247
Contribuição para o diferencial devido a:				
Ensino fundamental completo	1.73	-3.03 **	-3.44 **	0.11
Ensino médio incompleto	-9.74 **	-10.25 **	-11.47 **	-10.01 **
Ensino médio completo	2.08	20.82 **	18.61 **	18.20 **
Ensino superior incompleto	18.34 **	19.20 **	19.07 **	17.56 **
Ensino superior completo	76.51 **	71.71 **	65.67 **	65.49 **
Experiência	-1.15 **	1.53 **	0.32	0.83 **
Branco	-6.84 **	-17.80 **	-12.71 **	-13.93 **
Migrante	0.33 **	-0.04	0.14	0.07
Mulher	4.85 **	4.03 **	4.88 **	4.32 **
Casamento	0.06	0.17	0.09	0.04
Presença de filhos	0.09	-1.35 **	-0.65	-0.20
Região metropolitana	0.13	0.40 **	0.39 **	0.31 **
Área rural	-1.44	-1.14	-1.16	-1.61 **
Sindicalizado	-0.01	-0.10	-0.14	-0.10
Tipo de ocupação	-7.35 **	-3.88 **	-5.09 **	-4.21 **
Ramo de atividade	6.21 **	-9.47 **	2.04	-1.77
Total explicado (%)	83.8	70.82	76.56	75.08

Fonte: Elaboração Própria dos Autores a partir dos Microdados da PNAD de 2007.

Notas: ** $p \leq 0,05$. As contribuições estimadas reportadas são os valores médios obtidos para as decomposições usando 100 subamostras da região Sudeste. Regressões ajustadas para experiência (experiência e experiência ao quadrado), presença de filhos (menores de 06 anos e entre 07 e 14 anos), tipos de ocupação (carteira assinada e servidor público), ramos de atividade (Setor Agrícola e Extrativo; Indústrias de Transformação; Construção; Comércio e Serviços Pessoais; Serviços Financeiros etc.; Administração Pública e Serviços Sociais; Outros Serviços) e macrorregiões brasileiras (não computadas na decomposição acima).

incidência de sobre-educação 16,4% superior na região Sudeste, em comparação ao Nordeste. De acordo com os números da referida tabela, deste diferencial entre as regiões, as variáveis observáveis incluídas no modelo teriam um papel relevante no entendimento deste diferencial. As características observáveis consideradas explicariam entre 70,8 e

83,8% do diferencial de sobre-educados observado entre as regiões.

Diferenças nas condições de oferta de trabalho também se mostram relevantes na determinação do diferencial de sobre-educação entre o Sudeste e o Nordeste. Novamente, a distribuição do nível de

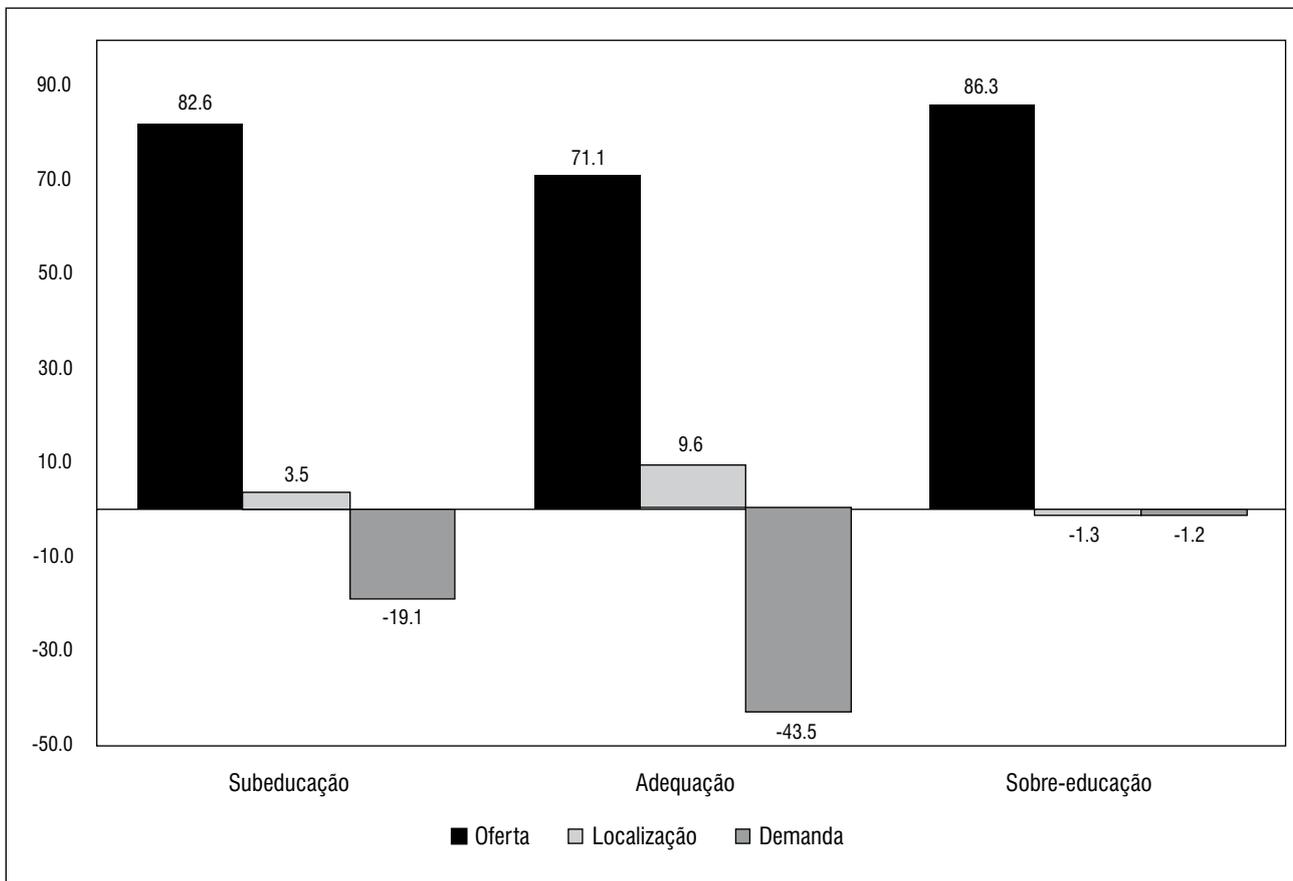


Gráfico 2 – Determinantes das Diferenças Regionais de *Mismatch* – Fatores de Oferta e Demanda de Trabalho.

Fonte: Elaboração Própria dos Autores a partir dos Dados da PNAD de 2007.

educação dos indivíduos entre as regiões responderia por grande parte das condições de oferta que explicam o *mismatch*. Uma proporção menor de indivíduos com ensino superior incompleto ou completo na região Nordeste responderia, respectivamente, por até 19,2% e 76,5% do diferencial de sobre-educação entre as regiões. Ao passo que, se o Sudeste detivesse a mesma proporção de indivíduos com ensino médio incompleto, o diferencial seria entre 9,7 e 11,5% menor que o observado. Logo, a concentração maior de indivíduos nos níveis mais elevados de educação na região Sudeste, comparativamente à região Nordeste, é fundamental para compreendermos o diferencial verificado.

Os resultados estimados para contribuição da experiência no mercado de trabalho não são robustos, variando a direção do efeito de acordo com amostra utilizada para estimação do modelo *probit*. A

distribuição de indivíduos que se declararam de cor/raça contribui para redução do diferencial de sobre-educação. A maior concentração de indivíduos de cor/raça branca no Sudeste, em relação ao Nordeste, associada à menor probabilidade de estes indivíduos serem classificados de sobre-educados, reduz o diferencial esperado entre as regiões. A participação das mulheres nas amostras (um pouco superior na região Nordeste) impacta, positivamente, entre 4,0 e 4,9% para o diferencial de sobre-educação. Do mesmo modo, a distribuição dos indivíduos entre as regiões metropolitanas explicaria uma pequena parcela do diferencial de sobre-educados entre as duas regiões (de 0,3 a 0,4%). Adicionalmente, a migração, casamento, áreas rurais e presença de filhos não contribuiriam estatisticamente para a explicação do diferencial de sobre-educação entre as regiões Nordeste e Sudeste.

Fatores de demanda de trabalho também parecem influir sobre o diferencial regional esperado de sobre-educação. A distribuição da mão-de-obra entre os diversos tipos de ocupação também explica parte do diferencial de sobre-educação entre as regiões. A concentração maior de indivíduos em tipos de ocupação informais (trabalho sem carteira assinada e conta própria) na região Nordeste, em relação ao Sudeste, que oferecem um risco maior de sobre-educação, eleva a taxa de sobre-educação esperada e ajuda a reduzir em até -7,4% do diferencial observado. Entretanto, a contribuição da distribuição dos ocupados entre os ramos de atividade nas duas macrorregiões não apresenta resultados robustos.

Num esforço de síntese dos resultados apresentados nas Tabelas 6, 7 e 8, o Gráfico 2, a seguir, apresenta resumidamente os efeitos dos fatores da demanda e oferta de trabalho para as condições sobre-educação, subeducação e os adequadamente alocados.

De forma geral, observando-se os valores para a condição de adequação, nota-se que os fatores da oferta contribuem para ampliar os diferenciais regionais (71,1% da diferença observada entre o NE e SE), influência que é auxiliada pelas variáveis de localização (região metropolitana e área rural/urbana, com cerca de 9,5% desta diferença), ao passo que os fatores da demanda atuam no sentido de reduzir (-43,5% de tal diferença entre as regiões) os diferenciais regionais. Mais especificamente para o caso dos diferenciais regionais de subeducação e sobre-educação entre tais regiões, percebe-se que os fatores de oferta de trabalho são responsáveis, respectivamente, por cerca de 83% do diferencial de subeducação entre o NE e o SE e por cerca de 86% do diferencial de sobre-educação entre o SE e o NE. Por outro lado, as estruturas ocupacionais associados à demanda de trabalho (tipos de ocupação e ramos de atividade) são relevantes para atenuar a posição desfavorável do NE em termos de subeducação (responsáveis que são por -19% do diferencial regional), mas pouco contribuem para atenuar a diferenças entre o SE e o NE em termos de níveis de sobre-educação. Finalmente, nota-se que, consideradas as características individuais e de tipos

de ocupações, as variáveis associadas à localização pouco explicam dos diferenciais de *mismatch* observados.

6 – CONCLUSÕES

A inadequação do nível de educação dos indivíduos no mercado de trabalho aferida a partir de microdados, apesar de ser discutida na literatura internacional, somente nos últimos anos começa a ser estudada para o Brasil frente à elevação dos níveis educacionais em seus diferentes graus. Santos (2002) e Diaz e Machado (2008) apontam para uma grande discrepância entre os níveis de inadequação observados nas regiões Nordeste e Sudeste do Brasil sem, contudo, levantarem qualquer hipótese conclusiva para o fenômeno observado. Este trabalho apresenta um primeiro esforço no sentido de apontar potenciais determinantes do nível de inadequação dos indivíduos no mercado brasileiro e regional e aferir a contribuição de diversas características observáveis relacionadas às condições de oferta e demanda por trabalho sobre o diferencial observado entre as regiões Nordeste e Sudeste.

Inicialmente, no que se refere aos níveis de *mismatch* ou desajustamento dos mercados de trabalho nacional e regionais e utilizando informações mais recentes, verificou-se, mais uma vez, a presença de significativas disparidades regionais: enquanto, para o Brasil e para a região SE, os níveis de *mismatch* giram em torno de 50% do pessoal ocupado, para o NE, tal nível de desajustamento está acima deste percentual. Verificou-se que particularmente elevado é o diferencial de subeducação (*undereducation*) entre as regiões NE e SE: enquanto a primeira região apresenta cerca de 33% do seu pessoal ocupado com níveis de escolaridade abaixo do requerido para as ocupações, no SE, tal percentual gira em torno de 25%.

Com respeito aos potenciais determinantes de *mismatch* dos mercados de trabalho brasileiros, as evidências obtidas indicam que tanto variáveis de oferta de trabalho, principalmente o término dos diferentes ciclos de escolaridade, como variáveis de demanda, preponderantemente os ramos de atividades, estão associados aos desajustamentos dos mercados de trabalho nacionais e regionais. Notou-se,

particularmente, que o término do ensino fundamental, médio ou superior (efeito-diploma), relativamente às situações de incompletude dos referido ciclos, diminui as chances de os ocupados estarem na situação subeducado, um efeito que não está presente no caso das chances de sobre-educação. Mais particularmente, mostrou-se que, enquanto o sexo feminino e a residência nas regiões metropolitanas são situações que diminuem as chances de os ocupados estarem na condição de subeducado, as situações de sindicalizado e residência no meio rural aumentam tais chances.

Mais importante no que tange às disparidades regionais no país, notou-se que, para a condição de subeducação, a região Nordeste teria um percentual de *mismatch* até 21,2% superior àquele observado na região Sudeste. Deste diferencial, diferenças nas características individuais entre as macrorregiões explicariam de 54,5% a 66,9% do diferencial regional. Novamente, os níveis educacionais divergentes entre as regiões seriam importantes para explicar o diferencial de *mismatch* observado, particularmente para os níveis de fundamental e superior completos. Embora com menor relevância, a maior concentração dos indivíduos nas áreas rurais no Nordeste também contribuiria para elevar o diferencial (3% do diferencial). Por outro lado, considerando fatores de demanda de trabalho, os diferenciais regionais quanto aos tipos de ocupação e à distribuição dos indivíduos entre os ramos de atividade nas duas regiões também seriam importantes na explicação do diferencial de subeducação, embora atuem no sentido de reduzir as disparidades regionais de subeducação: cerca de 19% do diferencial seria reduzido, caso as regiões apresentassem as mesmas estruturas de demanda. Assim, as condições de oferta de trabalho ampliariam o diferencial, ao passo que as condições de demanda de trabalho o reduziriam.

Finalmente, os resultados indicam que a região Sudeste teria uma incidência de sobre-educação até 16,3% maior que aquela observada no Nordeste. Este diferencial seria quase totalmente explicado pelas variáveis observáveis (83,8%). Deste diferencial, as condições de oferta de trabalho seriam as mais relevantes nesta explicação (responsáveis por 86% da disparidade), sendo pouco relevantes as condições de demanda de trabalho.

Uma vez que, como mostraram Diaz e Machado (2008), à condição de subeducado está associada uma penalidade salarial (para a mesma escolaridade, ocupados nesta condição têm remuneração abaixo daquela recebida pelos ocupados adequadamente alocados), de forma geral, os resultados obtidos nesta investigação indicam que os baixos níveis de escolaridade do trabalhador nordestino impõem-lhe uma dupla penalização: além da baixa remuneração associada por si à baixa escolaridade (mesmo quando adequadamente alocado), esta baixa escolaridade por vezes impede um melhor ajustamento deste às exigências de qualificação das ocupações, criando uma penalidade adicional. Se o objetivo é elevar o bem-estar da população, as políticas de desenvolvimento regional do Brasil devem, pois, considerar não apenas a formação generalizada de capital humano no NE, mas também considerar o grau de ajustamento da mão-de-obra local às condições de demanda presente e futura.

ABSTRACT

Based on the results recently pointed out by Santos (2002) e Diaz e Machado (2008) about the Brazilian regional levels of labor market mismatch and by using recent PNAD (Pesquisa Nacional por Amostra de Domicílios of 2007) microdata, the paper investigates the determinants of these regional labor market imperfect adjustments, and based on the decomposition proposed by Fairlie (2003) for non linear estimates, the factors explain the worst situation of the Northeast region, especially in what regards to the undereducated conditions. The evidences obtained indicate that both job offer (mainly education levels) and demand (sectors of activities, for example) affect the levels of mismatch in Brazilian regional labor markets. Both sets of factors also have an important role to play in explaining the highest level of mismatch of Northeast region when compared to Southeast region.

KEY WORDS

Labor Market. Overeducation. Undereducation. Regional Disparities.

REFERÊNCIAS

BAUER, T. Educational mismatch and wages: a panel analysis. **Economics of Education Review**, v. 21, p. 221-229, 2002.

BRASIL. Ministério do Trabalho e Emprego. **Classificação Brasileira de Ocupações: CBO 2002**. Brasília, DF, [20--]. Disponível em: <www.mtecbo.gov.br>. Acesso em: 6 abr. 2009.

CAHUC, P.; ZYLBERBERG, A. **Labor economics**. Cambridge: Massachusetts Institute of Technology, 2004.

DIAZ, M. D. M.; MACHADO, L. Overeducation e undereducation no Brasil: incidência e retornos. **Estudos Econômicos**, v. 38, n. 3, p. 431-460, 2008.

DOLTON, P.; VIGNOLES, A. The incidence and effects of overeducation in the UK graduate labour market. **Economics of Education Review**, v. 19, n. 2, p. 179-198, 2000.

DUCAN, G.; HOFFMAN, S. The incidence and wage effects of overeducation. **Economics of Education Review**, v. 1, p. 75-86, 1981.

FAIRLIE, R. W. The absence of the African-American owned business: an analysis of the dynamics of self-employment. **Journal of Labor Economics**, v. 17, n. 1, p. 80-108, 1999.

_____. **An extension of the Blinder-Oaxaca decomposition technique to logit and probit models**. Yale: Yale University, 2003. (Economic Growth Center. Discussion paper, n. 873).

GROOT, W. The incidence of, and returns to overeducation in the UK. **Applied Economics**, v. 28, n. 10, p. 1345-1350, 1996.

GROOT, W.; MAASSEN VAN DEN BRINK, H. Overeducation in the labor market: a

meta-analysis. **Economics of Education Review**, v. 19, p. 149-158, 2000.

HARTOG, J. Over-education and earnings: where are we, where should we go?. **Economics of Education Review**, v. 19, p. 131-147, 2000.

HARTOG, J.; TSANG, M. **Education, job level and earnings in the US, 1969-1973-1977**. Amsterdam: University of Amsterdam, 1989.

IBGE. **Censo demográfico 2000: população: domicílios**. Rio de Janeiro, 2000. CD-ROM.

_____. **Pesquisa Nacional por Amostra Domiciliar**. [S.l.], 2007. CD ROM.

LINSLEY, I. **Causes of overeducation in the Australian labour market**. Melbourne: University of Melbourne, 2005. (Research Paper, n. 940).

MACHADO, A. F.; OLIVEIRA, A. C. de; CARVALHO, N. F. **Tipologia de qualificação da força de trabalho: uma proposta a partir da noção de incompatibilidade entre ocupação e escolaridade**. Belo Horizonte: Universidade Federal de Minas Gerais, 2003. (Texto para Discussão, n. 218).

MCGOLDRICK, K.; ROBST, J. Gender differences in overeducation: a test of the theory of differential overqualification. **The American Economic Review**, v. 86, p. 280, 1996.

MCGUINNESS, S. Overeducation in the labour market. **Journal of Economic Surveys**, v. 20, n. 3, p. 387-418, 2006.

NEUMARK, D. Employers' discriminatory behavior and the estimation of wage discrimination. **Journal of Human Resources**, v. 23, p. 279-295, 1988.

OAXACA, R.; RANSOM, M. On discrimination and the decomposition of wage differentials. **Journal of Econometrics**, v. 61, p. 5-21, 1994.

- RUMBERGER, R. The rising incidence of overeducation in the US labour market. **Economics of Education Review**, v. 1, p. 293-314, 1981.
- SANTOS, A. M. dos. Overeducation no mercado de trabalho brasileiro. **Revista de Economia de Empresas**, v. 2, n. 2, p. 1-22, 2002.
- SANTOS JÚNIOR, E. R.; MENEZES-FILHO, N.; FERREIRA, P. C. Migração, deleção e diferenças regionais de renda no Brasil. **Pesquisa e Planejamento Econômico**, v. 35, n. 3, p. 299-331, 2005.
- SICHERMAN, N. 'Overeducation' in the labour market. **Journal of Labour Economics**, v. 9, p. 101-122, 1991.
- SILVA, T. F. B.; SILVEIRA NETO, R. M. Migração e seleção no Brasil: evidências para o decênio 1993-2003. In: ENCONTRO DE ECONOMIA REGIONAL DO NORDESTE, 10., 2005, Fortaleza. **Anais...** Fortaleza: Banco do Nordeste do Brasil, 2005.
- SPENCE, M. Job market signaling. **The Quarterly Journal of Economics**, v. 87, n. 3, p. 355-374, Aug. 1973.
- VERDUGO, R.; VERDUGO, N. The impact of surplus schooling on earnings: some additional findings. **Journal of Human Resources**, v. 24, p. 629-643, 1989.

Recebido para publicação em: 19.07.2010.