

Retorno Salarial do *Overeducation*: Viés de Seleção ou Penalização ao Excesso de Escolaridade?

ANA CLÁUDIA ANNEGUES*

WALLACE PATRICK SANTOS DE FARIAS SOUZA†

Sumário

1. Introdução	119
2. <i>Overeducation</i> : Teoria e Evidência Empírica	122
3. Estratégia Empírica	124
4. Dados	127
5. Resultados	131
6. Considerações Finais	135

Palavras-chave

Overeducation, salário, viés de seleção

JEL Codes

I20, J31, C23


Resumo • Abstract


Este paper investiga os efeitos salariais do *overeducation*, situação na qual trabalhadores exercem funções com escolaridade requerida inferior. Para isso, utilizou-se dados de graduados da Universidade Federal da Paraíba de 2003 a 2013. A estimação de um OLS com efeitos fixos dos indivíduos e das firmas, mais interações destes com o tempo como forma de captar os fatores não observados *time varying* mostrou que egressos *overeducated* auferem em média um salário menor que seus pares com bom *match* entre instrução e ocupação. A regressão quantílica mostrou que este efeito penaliza em maior grau os indivíduos nos quantis superiores da distribuição de salários, enfraquecendo a hipótese que liga o *overeducation* a menores níveis de habilidade.

1. Introdução

A sobreeducação, em inglês *overeducation*, é o fenômeno no qual trabalhadores desempenham funções que exigem um grau de instrução inferior à possuída por estes. Na sua origem está a expansão da oferta de trabalhadores qualificados, grande parte com ensino superior, não acompanhada pelo crescimento em mesma proporção da demanda no mercado de trabalho, gerando perfis de trabalhadores além ou aquém do requerido.

Nas últimas décadas a literatura tem se empenhado em compreender as causas e as consequências do *overeducation*, embora não apresente uma teoria unânime sobre

*Pós-doutoranda do Programa de Pós-Graduação em Economia do Centro de Ciências Sociais Aplicadas, Universidade Federal do Rio Grande do Norte (PPECO/CCSA/UFRGN). Av. Senador Salgado Filho, 3000, Lagoa Nova, Natal, RN, CEP 59072-970, Brasil.  0000-0001-6656-3824

†Programa de Pós-Graduação em Economia, Departamento de Economia, Universidade Federal da Paraíba (PPGE/DE/UFPB). Via Expressa Padre Zé, 289, Castelo Branco III, João Pessoa, PB, Brasil.  0000-0002-0598-936X

✉ annegues.ana@gmail.com ✉ wpsfarias@gmail.com

o tema. Segundo Groot (1996), dentre as consequências do *overeducation* estão a subutilização do estoque de capital humano e, conseqüentemente, a ineficiência dos gastos públicos em educação

Parte significativa das pesquisas empíricas existentes se dedica a analisar as consequências em termos salariais para os trabalhadores *overeducated* (como Bauer, 2002, Sicherman, 1991, and McGuinness, 2006). Em geral os resultados encontrados apontam para uma redução nos rendimentos desses trabalhadores com relação aos seus pares corretamente alocados no mercado.

O procedimento adotado por estes trabalhos consiste em estimar uma equação minceriana (Mincer, 1974) de salários acrescentando o excesso de anos de estudos com relação ao que é requerido ou uma variável *dummy* indicando se o indivíduo está ou não em *overeducation*. Entretanto, tais abordagens empíricas enfrentam o mesmo desafio imposto às análises tradicionais de retorno da educação: o controle da heterogeneidade não observada entre os indivíduos. Isso implica em identificar se os trabalhadores *overeducated* auferem um salário menor pelo excesso de qualificação ou por fatores não observados que levam o trabalhador a viver esse status. Sendo assim, são necessárias estratégias que lidem com esse viés de seleção.

Boa parte das pesquisas iniciais sobre o impacto salarial do *overeducation* utilizam apenas dados em *cross-section*, o que dificulta o controle da heterogeneidade não observada e pode resultar em estimativas tendenciosas caso haja correlação entre habilidades inatas e *overeducation*. Chevalier (2003), embora utilize dados de corte, tenta incluir as habilidades não observadas explicitamente no modelo utilizando os resíduos da equação minceriana tradicional como *proxy*. Já Bauer (2002) procura fazer esse controle com dados em painel e estimação com efeitos fixos, o que permite o controle de características não observadas fixas no tempo.

No entanto, o viés surge também da existência de não observáveis variantes no tempo e somente o controle dos efeitos fixos pode ser insuficiente para fornecer estimativas robustas. A existência de habilidades que mudam com o tempo é uma hipótese razoável, principalmente se considerarmos o argumento levantado por alguns autores (Duncan & Hoffman, 1981; Sicherman, 1991; Cohn & Ng, 2000) de que o *overeducation* pode ser uma maneira informal de aquisição de capital humano; através da vivência do mercado de trabalho, o indivíduo pode adquirir habilidades que não são repassadas nos cursos universitários. Uma alternativa para controlar esses fatores seria encontrar uma variável instrumental para o *overeducation*, o que não é uma tarefa trivial.

McGuinness (2006) afirma que as metodologias de Chevalier (2003) e Bauer (2002) ainda ignoram os potenciais impactos salariais de componentes não observados. Para testar se o *overeducation* se origina de um déficit de habilidade por parte dos indivíduos, o autor parte da hipótese de que a localização do indivíduo na distribuição de salários reflete o seu nível de habilidade relativa, de modo que indivíduos mais habilidosos tendem a ocupar posições mais elevadas na distribuição. A equação minceriana é estimada, então, via Regressão Quantílica; assim, é possível quantificar a redução de salário que trabalhadores *overeducated* sofrem com relação aos corretamente alocados com mesmo nível de habilidade. Se o *overeducation* reduz mais fortemente o salário dos trabalhadores menos habilidosos, é um indício de que as habilidades não observadas

desempenham um papel importante na situação de excesso de qualificação desses indivíduos.

Dito isso, o presente artigo objetiva mensurar o impacto do *overeducation* sobre os salários dos indivíduos com ensino superior empregando ferramentas para contornar os problemas de autosseleção. Consideramos um modelo empírico com especificação semelhante à equação minceriana de salários, incluindo uma *dummy* para os indivíduos *overeducated*. Primeiramente, procedemos à estimação de um modelo de dados em painel com dois efeitos fixos, relacionados ao trabalhador e à firma na qual o mesmo trabalha. Aqui queremos incluir também características não observadas das empresas que podem impactar o salário.

Dada a dificuldade de encontrar instrumentos adequados para o *overeducation*, a estratégia adotada para controlar a parcela dos fatores omitidos que são variantes no tempo é inspirada no trabalho de Baltagi, Egger, e Pfaffermayr (2014) da literatura de comércio internacional a qual consiste na inclusão de *dummies* para interações entre trabalhador e ano e interações entre empresa e ano. A utilização da dimensão de tempo fornece a variação de características não observadas dos trabalhadores e das firmas entre os anos, proporcionando uma aproximação das não-observáveis *time-varying*. A estimação do modelo é feita via *Ordinary Least Squares* (OLS) para dados em painel com mais de duas dimensões. Esse método foi desenvolvido por Carneiro, Guimarães, e Portugal (2009) e aperfeiçoado computacionalmente por Guimarães e Portugal (2009).

Mesmo com as interações como *proxies* de fatores variantes no tempo, queremos ainda investigar quais os papéis da habilidade não observada e do *overeducation* na determinação dos salários. Para tanto, tal qual McGuinness (2006), estimamos a equação minceriana por Regressão Quantílica. Caso o *overeducation* reduza os salários dos trabalhadores em diferentes quantis em proporção semelhante, pode-se lançar dúvida sobre a hipótese que liga o *overeducation* exclusivamente às diferenças de habilidade. Além disso, a regressão quantílica permite visualizar os diferentes impactos do excesso de qualificação em diferentes níveis de salários.

Serão utilizados dados de egressos da Universidade Federal da Paraíba (UFPB), fornecidos pela Superintendência de Tecnologia da Informação (STI) na universidade, entre os anos de 2003 e 2013. Através do Cadastro de Pessoa Física (CPF), foi possível mapear a situação desses ex-alunos na Relação Anual de Informações Sociais (RAIS), identificando as suas ocupações em cada ano, segundo a Classificação Brasileira de Ocupações de 2002 (CBO 2002). Na CBO são definidos os graus de instrução necessários a cada ocupação, o que permite saber se o egresso está ou não *overeducated*.

Os dados foram agrupados em seis janelas de seis anos (2003–2008, 2004–2009, 2005–2010, 2006–2011, 2007–2012, 2008–2013).¹ Essa divisão foi feita para evitar a perda de observações que ocorreria caso analisássemos somente os indivíduos permanentes na amostra durante todo o período 2003–2013. A saída do indivíduo (atrito), se ocorrida de forma não aleatória, pode implicar em mais uma fonte de heterogeneidade não observada.

O artigo encontra-se dividido da seguinte forma: a seção 2 contém uma revisão teórica e da literatura empírica do *overeducation*; na seção 3 tem-se os procedimentos

¹Procedimento semelhante foi empregado por Schank, Schnabel, e Wagner (2010), que investigaram a existência de viés de seleção nos maiores salários pagos por firmas exportadoras

metodológicos; na seção 4 a descrição da base de dados; na seção 5 os resultados encontrados; e na seção 6 as considerações finais.

2. *Overeducation*: Teoria e Evidência Empírica

Até o momento, a literatura tem interpretado o *overeducation* sob a ótica de algumas teorias sobre educação e mercado de trabalho, como a teoria do capital humano (Becker, 1962, e Schultz, 1961), os modelos de competição por emprego (Thurow, 1975) e os assignment models.

Segundo a teoria do capital humano (ver Becker, 1962, e Schultz, 1961), os salários dos trabalhadores são determinados pela sua produtividade, a qual depende do seu estoque de capital humano adquirido pela educação formal e pela experiência. Nesse sentido, o *overeducation* pode ser considerado uma exceção à teoria do capital humano por originar-se de um desequilíbrio entre oferta e demanda no mercado de trabalho. A consistência da teoria do capital humano, frente à ocorrência de *overeducation*, seria possível caso o *overeducation* fosse comprovadamente um fenômeno de curto prazo, onde o mercado naturalmente se encarregaria de corrigir os desajustes entre trabalhadores e firmas.

Sob a perspectiva dos modelos de competição por emprego (Thurow, 1975), o excesso de educação formal seria uma consequência da competição entre os trabalhadores pelos postos de trabalho. A rigidez da demanda por trabalhadores qualificados levaria estes a acumularem mais anos de estudo, em alguns casos mais do que o necessário para uma determinada ocupação. Embora o aumento da oferta de trabalhadores qualificados reduza o retorno à escolaridade, o investimento em educação não diminui, dado que este assegura a posição do trabalhador na competição pela vaga de trabalho. Além disso, os trabalhadores podem permanecer em empregos para os quais são *overeducated* com vistas a adquirir a habilidade necessária para conseguir uma ocupação adequada ao seu grau de instrução (Sicherman, 1991). Nessas condições, o *overeducation* tende a se tornar uma condição permanente aos indivíduos, contrapondo a tese levantada pela teoria do capital humano de que se trataria de um fenômeno de curto prazo.

Uma explicação alternativa, a qual seria um meio termo entre as duas hipóteses levantadas anteriormente, é fornecida pelos *assignment models*. No geral, esses modelos pontuam que a distribuição de salários entre os trabalhadores é um reflexo do processo de alocação de trabalhadores heterogêneos à diferentes tipos de ocupações. Sattinger (1993) mostra que essa desigualdade veio se reforçando ao longo do tempo e a explicação para isso está relacionada não só à qualificação, mas também depende da disponibilidade e/ou qualidade dos empregos (McGuinness, 2006). A partir do momento que há um desencontro entre a distribuição de empregos segundo às exigências de habilidade e a distribuição de trabalhadores segundo os níveis de educação, alguns trabalhadores acabam em trabalhos para os quais estão sobreeducados (Kiersztyn, 2013).

Os primeiros estudos empíricos a respeito do descasamento entre grau de instrução possuído e grau de instrução requerido surgem nos anos 70 nos Estados Unidos, com os estudos de Freeman (1976) e Smith e Welch (1978), quando se torna iminente a preocupação com o excesso de oferta de trabalhadores qualificados daquele país.

Dentre os determinantes micro e macroeconômicos do *overeducation* comumente analisados pela literatura se destacam as características individuais, como idade e gênero, as condições do mercado de trabalho e a taxa de desemprego da economia. [Morano \(2014\)](#) testa a influência de inúmeros desses determinantes sobre a probabilidade de um trabalhador se encontrar em uma situação de *overeducation*, utilizando dados do mercado de trabalho da Itália. Com relação à idade, os autores encontram que os trabalhadores mais jovens são mais propensos à ocupar postos de trabalho não condizentes com seu grau de instrução. Este resultado confirma o consenso geral existente na literatura de que a qualidade do ajuste entre instrução e ocupação aumenta com a idade do indivíduo. Algumas teorias sobre mobilidade na carreira preveem que as habilidades adquiridas durante a situação de *overeducation* podem aumentar a probabilidade de ser promovido, de modo que ser *overeducated* seria uma estratégia ótima no início da carreira ([Leuven & Oosterbeek, 2011](#)).

Além dos atributos individuais e das condições macroeconômicas, uma parte importante da literatura tem sugerido que o *overeducation* consiste em um fenômeno mais comum em determinadas áreas de formação. As áreas ligadas a ciências e tecnologia podem fornecer maiores níveis de “habilidades produtivas” do que as áreas de humanidades, pois o currículo desses cursos enfatizam mais a aquisição de habilidades específicas requeridas pelas ocupações no mercado de trabalho ([Reimer, Noelke, & Kucel, 2008](#)).

Tal hipótese vem sendo corroborada pelas evidências empíricas. [Dolton e Vignoles \(2000\)](#) e [Dolton e Silles \(2008\)](#) analisaram os determinantes do *overeducation* no mercado de trabalho entre indivíduos graduados no Reino Unido. Ambos os trabalhos encontram que trabalhadores formados nas áreas de humanidades e artes apresentaram maior probabilidade de estarem sobreeducados em comparação com outros campos de estudo

[Duncan e Hoffman \(1981\)](#) foram os pioneiros a comparar os anos de estudos possuídos por um trabalhador e os requeridos pela sua ocupação, e tentaram estimar os impactos dessa diferença nos rendimentos do trabalhador. Para isso, os autores estimaram uma equação minceriana, distinguindo no modelo de determinação de salários os anos de estudos possuídos e requeridos, e como resultado encontram que a educação requerida possui um retorno salarial duas vezes maior que a educação possuída pelo indivíduo. Desde então, parte da literatura tem utilizado estratégias semelhantes para medir os impactos salariais da sobreeducação. Muitos desses estudos apontam que indivíduos *overeducated* possuem um retorno salarial menor que os trabalhadores cujos empregos correspondem ao nível de instrução ([Rumberger, 1981](#); [Sicherman, 1991](#); [Groot, 1996](#)).

No Brasil, há ainda um número reduzido de pesquisas sobre o *overeducation*. [Santos \(2002\)](#) procurou verificar a incidência e o prêmio para sobreeducação nos anos de 1992, 1995, 1997 e 1999, a partir dos dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) e encontram uma penalização de 21,25% para os trabalhadores *overeducated*. Já [Diaz e Machado \(2008\)](#) utilizando dados do Censo 2000, encontram uma maior incidência do *overeducation* na região Sul, com um retorno salarial de 12,21%.

Recentemente, [Annegues, Oliveira, Figueiredo, e Porto \(2018\)](#), também utilizando a base de dados de egressos da Universidade Federal da Paraíba (UFPB) cruzada com dados sobre mercado de trabalho da Relação Anual de Informações Sociais (RAIS) investigou quais fatores afetam a probabilidade de um indivíduo estar em uma ocupação com nível de instrução exigido abaixo da sua formação superior, em especial o efeito da área de formação dos egressos. As áreas de humanidades, artes e sociais aplicadas apresentam maior probabilidade dos indivíduos estarem sob *overeducation*, bem como uma menor chance de deixar essa condição durante alguns anos após a conclusão do curso. [Oliveira, Mariano, e Araújo \(2019\)](#), com dados de indivíduos do Ceará, mostraram que a expansão do ensino superior verificada nos últimos anos aumentou a probabilidade de um egresso se encontrar em situação de *overeducation*.

3. Estratégia Empírica

3.1 Modelo Empírico

O modelo comumente estimado pela literatura consiste em uma versão modificada da equação de salários minceriana, com especificações inspiradas no trabalho de [Verdugo e Verdugo \(1989\)](#):

$$\ln w_i = \beta_0 + \beta_1 E_i + \beta_2 \exp_i + \beta_3 \exp_i^2 + \beta_4 OV_i + \varepsilon_i, \quad (1)$$

onde o *overeducation* é medido por uma *dummy* com valor 1 se o trabalhador i está em uma ocupação com escolaridade requerida inferior a que ele possui. Os trabalhos que utilizam a especificação (1) se baseiam em dados de cross-section para estimar os coeficientes. Entretanto, a possível existência de variáveis omitidas no modelo, como as habilidades inatas, que determinam tanto o salário quanto a condição de *overeducation*, pode comprometer a validade dessas estimativas. Uma forma de contornar esse problema é o controle da heterogeneidade não observada com a aplicação de técnicas de estimação utilizando dados em painel. O modelo de efeitos fixos consiste em introduzir uma *proxy* para as habilidades individuais e demais fatores fixos no tempo.

Já o controle de não-observáveis variantes no tempo requer o emprego de outras técnicas de identificação, como a abordagem de variáveis instrumentais. Encontrar um bom instrumento, o qual deve ser correlacionado com a variável explicativa endógena e não correlacionado com o termo de erro na equação, nem sempre é uma tarefa trivial, tendo em conta principalmente as limitações impostas pela disponibilidade de dados. Assim, com o objetivo de tratar a presença de fatores fixos e variantes no tempo, o modelo que estimamos para o retorno salarial do *overeducation* inclui efeitos fixos do empregado, da firma, e interações destes com os efeitos de tempo. A especificação (1) assume a forma abaixo:

$$\ln w_{ijt} = \beta_0 + \beta_1 E_{it} + \beta_2 \exp_{it} + \beta_3 \exp_{it}^2 + \beta_4 OV_{it} + \alpha_i + \delta_j + \alpha_{it} + \delta_{jt} + janela_{it} + \varepsilon_{ijt} \quad (2)$$

onde o *overeducation* é representado pela *dummy* OV_{it} , que assume valor igual a 1 para os trabalhadores que estão *overeducated* em um determinado ano t ; $\ln w_{ijt}$ é o logaritmo do salário do indivíduo i que trabalha na empresa j no ano t ; E_{it} os anos de estudo adquiridos; e \exp_{it} é a experiência.

Diferentemente das abordagens de painel existentes para o efeito do *overeducation* sobre salário, aqui introduzimos um efeito fixo do indivíduo e um efeito fixo relacionado à firma, representados na equação (2) pelos termos α_i e δ_j , respectivamente. Além de determinar o salário do empregado, características da empresa podem determinar a contratação do indivíduo superqualificado para a vaga de trabalho. Segundo Santos (2002), a existência de *overeducation* no Brasil sinaliza a carência de mão-de-obra qualificada, de maneira que as empresas, na busca de trabalhadores produtivos, vêm na escolaridade mais elevada desses trabalhadores uma sinalização de maior produtividade, contratando-os para os postos.

As interações dos efeitos dos trabalhadores e das firmas com o tempo surgem como uma aproximação das características não observadas variante no tempo. Esse procedimento é inspirado na estratégia de Baltagi et al. (2014). Os autores, analisando determinantes dos acordos comerciais, trabalham com uma estrutura de painel com efeitos em três dimensões: um efeito fixo relacionado aos pares de países com relações comerciais e interações de cada país com a dimensão de tempo. No modelo aqui estimado, as interações são representadas pelos termos α_{it} e δ_{jt} .

O termo $janela_t$ representa um conjunto de *dummies* para cada uma das seis janelas de seis anos na qual dividimos a amostra. A inclusão das janelas no modelo permite comparar as diferenças salariais entre os trabalhadores com excesso de qualificação e os corretamente alocados dentro de uma mesma janela de anos.

3.2 Estimação

O modelo de dados em painel a ser estimado possui mais de duas dimensões. Além do efeito fixo do indivíduo, tem-se os efeitos fixos da firma e os efeitos de interações indivíduo-ano e firma-ano. Isto demanda, portanto, um método de estimação para painéis com mais de duas dimensões. Carneiro et al. (2009) propuseram um método, empregado na estimação de um modelo de regressão com mais de 26 variáveis, um conjunto de dados de 26 milhões de observações e dois efeitos fixos, trabalhador e firma. A sua implementação, entretanto, se mostrou altamente custosa computacionalmente. Guimarães e Portugal (2009), então, desenvolveram um algoritmo que permite estimar um modelo com mais de um efeito fixo de forma mais simples.

Considere agora o seguinte modelo de regressão linear na sua forma matricial com a introdução de três efeitos fixos:

$$Y = Z\beta + D_1\lambda + D_2\theta + D_3\delta + \varepsilon, \quad (3)$$

em que Z é uma matriz de variáveis explicativas e D_1, D_2 e D_3 são matrizes com alta dimensão que comportam as *dummies* dos efeitos fixos. Assim, têm-se as equações matriciais

$$\begin{bmatrix} \beta = (Z'Z)^{-1}Z'(Y - D_1\lambda - D_2\theta - D_3\delta) \\ \lambda = (D_1'D_1)^{-1}D_1'(Y - Z\beta - D_2\theta - D_3\delta) \\ \theta = (D_2'D_2)^{-1}D_2'(Y - Z\beta - D_1\lambda - D_3\delta) \\ \delta = (D_3'D_3)^{-1}D_3'(Y - Z\beta - D_1\lambda - D_2\theta) \end{bmatrix},$$

as quais sugerem uma solução iterativa na qual a estimação de cada parâmetro do modelo depende dos demais. O ponto chave do procedimento de Guimarães e Portugal (2009) é que a dimensão das matrizes de variáveis *dummy* não mais representa

uma preocupação. Os termos $(D'D)^{-1}$ são matrizes genéricas que consistem em uma transformação dos dados com as médias intragrupos. Além disso, $D_1\lambda$, $D_2\theta$ e $D_3\delta$ entram nas equações como vetores coluna contendo todos os elementos de λ , θ e δ . O estimador usual de MQO considerando os efeitos fixos é fruto de uma regressão linear simples da variável dependente Y transformada (livre dos efeitos fixos) sobre o conjunto de variáveis exógenas Z . Aqui, em vez de transformar Y , ela é mantida como variável dependente em sua integridade e $D_1\lambda$, $D_2\theta$ e $D_3\delta$ entram na regressão como covariadas adicionais. Os coeficientes de $D_1\lambda$, $D_2\theta$ e $D_3\delta$ serão iguais a um e cada vetor por sua vez será formado pelos coeficientes estimados das variáveis *dummies* se estas fossem incluídas no modelo.

3.3 Regressão Quantílica

As estimativas do procedimento de [Guimarães e Portugal \(2009\)](#) fornecem o efeito médio do *overeducation*. No entanto, é possível que haja uma variabilidade de efeitos dentro da distribuição de salários e a estimação na média pode subestimar (superestimar) esses impactos.

Além disso, os valores estimados para cada quantil podem fornecer uma ideia do papel das habilidades individuais e do *overeducation* na determinação dos salários. A hipótese, pontuada por [McGuinness \(2006\)](#), assume que a posição relativa do indivíduo na distribuição dos rendimentos define seu grau de habilidade. Considerando que a amostra é composta por indivíduos com níveis similares de escolaridade, é razoável supor que indivíduos mais habilidosos estariam localizados nos quantis de salário mais altos. A regressão quantílica permite observar as diferenças salariais entre os *overeducated* e seus pares corretamente alocados dentro de um mesmo nível de habilidade; se, por exemplo, o *overeducation* penaliza em maior grau trabalhadores menos habilidosos, há um indício em favor da tese de que as diferenças de rendimento em razão do *overeducation* estão mais ligadas às diferenças de habilidades não observadas. Caso o *overeducation* impacte de forma semelhante trabalhadores com diferentes níveis de habilidade, põe-se em dúvida a tese da influência das habilidades.

Dito isso, também estimou-se o modelo descrito em (2) por regressões quantílicas. Outras vantagens da utilização de regressões quantílicas são: a robustez a *outliers*; o fornecimento de estimadores mais eficientes em comparação com os obtidos via OLS; e não necessidade da hipótese de distribuição normal.

O método de estimação que empregamos se trata de uma adaptação da abordagem de painel quantílico de [Canay \(2011\)](#). Originalmente, o método do autor consiste em uma transformação dos dados que consegue eliminar os efeitos fixos, considerando que estes afetam igualmente todos os quantis. Para uma compreensão mais detalhada da técnica, considere inicialmente o seguinte modelo de regressão em painel com efeito fixo:

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}, \quad \mathbb{E}(\varepsilon_{it} | x_{it}, \alpha_i) = 0, \quad (4)$$

onde $(y_{it}, x_{it}) \in \mathbb{R} \times \mathbb{R}^k$ são variáveis observáveis; $\theta(U_{it}) = \beta$ e $(U_{it}, \alpha_i) \in \mathbb{R} \times \mathbb{R}$ são fatores não observados; $U_{it} | x_{it}, \alpha_i \approx U[0, 1]$. Assume-se que a função $\tau \rightarrow x'\theta(\tau)$ é estritamente crescente em $\tau \in (0, 1)$ e que $\theta(\tau)$ representa o parâmetro de interesse.

A principal restrição é que as heterogeneidades individuais, representadas pelos efeitos fixos α_i , afetam a variável resposta dos indivíduos da mesma forma em todos os quantis.

Sendo $\mathbb{E}_T(\cdot) \equiv T^{-1} \sum_{t=1}^T (\cdot)$ e $\mathbb{E}_{nT}(\cdot) \equiv (nT)^{-1} \sum_{t=1}^T \sum_{i=1}^n (\cdot)$, Canay (2011) sugere um procedimento em dois estágios para estimação da equação (6).

Passo 1: Seja $\hat{\theta}_\mu$ um estimador \sqrt{nT} consistente de θ_μ . Os estimadores dos efeitos fixos têm a forma

$$\hat{\alpha}_i \equiv \mathbb{E}_T[y_{it} - x'_{it}\hat{\theta}_\mu]. \quad (5)$$

Passo 2: Fazendo a seguinte transformação na variável dependente $\hat{y}_{it} = y_{it} - \hat{\alpha}_i$, define-se o estimador de regressão quantílica $\hat{\theta}(\tau)$ como sendo

$$\hat{\theta}(\tau) = \arg \min_{\theta \in (\cdot)} \mathbb{E}_{nT}[\rho_\tau(\hat{y}_{it} - x'_{it}\theta_\mu)]. \quad (6)$$

No primeiro passo, procede-se a uma estimação OLS com as variáveis explicada e explicativas originais. Os resíduos da regressão contêm todos os fatores não observáveis que afetam y_{it} para cada indivíduo, de modo que a sua média no tempo, dada por $\mathbb{E}_T[y_{it} - x'_{it}\hat{\theta}_\mu]$, pode ser utilizada como uma *proxy* para os efeitos fixos individuais. Por fim, o segundo estágio consiste na aplicação do estimador de regressões quantílicas substituindo a variável dependente pelo seu valor deduzido dos efeitos fixos individuais.

Aqui adotamos um procedimento semelhante, porém não utilizamos a média dos resíduos da regressão OLS, mas sim os próprios efeitos fixos, cujos valores são fornecidos pelo método de Guimarães e Portugal (2009). O algoritmo desenvolvido pelos autores consegue salvar os efeitos fixos como variáveis, de todas as dimensões consideradas (emprego, firma e interações com o tempo). Em resumo, a) estimamos a equação (2) por OLS para mais de dois efeitos fixos, b) salvamos os efeitos fixos, c) construímos a variável salário livre dos efeitos fixos, e por fim d) estimamos a variável dependente transformada contra as demais variáveis explicativas através da regressão quantílica.

4. Dados

O banco de dados é resultado do cruzamento de informações acadêmicas dos egressos da Universidade Federal da Paraíba (UFPB), Campus I, localizada em João Pessoa, na Paraíba, fornecidas pela Superintendência de Tecnologia da Informação (STI) da UFPB e informações acerca da sua ocupação presentes na Relação Anual de Informações Sociais (RAIS). A variável de ligação entre as duas bases é o CPF do ex-aluno. Foram considerados os alunos graduados na UFPB a partir de 2003, acompanhados ao longo do tempo por pelo menos cinco anos. Assim temos as seguintes janelas de tempo: 2003–2008, 2004–2009, 2005–2010, 2006–2011, 2007–2012 e 2008–2013. Após a junção dos dados da UFPB com os dados da RAIS e a seleção dos egressos presentes em ambas as bases, tem-se um total de 4.598 observações.

A STI possui dados que abrangem a vida acadêmica de todos os alunos durante a graduação, como o semestre no qual ingressou e no qual finalizou o curso e a área de formação do concluinte (Ciências Jurídicas, Sociais Aplicadas, Saúde, Engenharia). Essa

informação é interessante, pois nos permite estimar o retorno salarial do *overeducation* em cada campo de atuação, tendo em vista as evidências de que o fenômeno tende a ser mais comum em determinados campos de estudo (ver seção 2).

Os dados ainda contêm informações coletadas na aplicação de um questionário socioeconômico no momento do vestibular, feita pela Comissão Permanente de Vestibular (COPERVE), responsável pela realização do processo seletivo para a UFPB até 2013. Os dados dizem respeito a características socioeconômicas dos alunos, como sexo, raça, grau de instrução dos pais (*background* familiar), e informações sobre sua vida escolar antes de entrar na faculdade (nota no vestibular, rede de ensino na qual cursou o ensino médio e etc.). Na RAIS constam as características individuais dos trabalhadores (raça, sexo, grau de instrução, idade, tipo de admissão, salário anual) e informações sobre as ocupações e sobre as empresas, como seu porte, classificação de atividade e natureza jurídica.

A construção do indicador de *overeducation* depende de como se mede a escolaridade requerida por uma determinada ocupação. Existem três formas principais nas quais a literatura se baseia: i) *Job Analysis*, ii) *Realized Matches*, e iii) *Self-Assessment*. A primeira consiste em utilizar uma classificação ocupacional, na qual analistas de trabalho definem os graus de escolaridade necessários para as ocupações. No *Realized Matches*, calcula-se a escolaridade média dos trabalhadores inseridos em cada ocupação e a que possuem escolaridade superior à média mais o desvio-padrão de sua ocupação são classificados como *overeducated*. Já no *Self-Assessment* o próprio trabalhador reporta o nível de instrução requerido pelo seu trabalho ou ele responde diretamente se o *match* entre sua ocupação e seu nível de instrução é adequado.

Dada a disponibilidade dos dados, aqui empregamos a abordagem do *Job Analysis* para definir se o egresso é ou não *overeducated*. Uma das variáveis presente na RAIS é a ocupação do egresso, segundo o sistema da Classificação Brasileira de Ocupações (CBO) de 2002. De acordo com o Ministério do Trabalho e Emprego (MTE) a ocupação é a agregação de empregos ou situações de trabalho similares quanto às atividades realizadas. Sendo assim, CBO 2002 agrega os empregos por habilidades cognitivas comuns exigidas no exercício de um campo de trabalho.²

O nível mais agregado da CBO 2002 são os grandes grupos formado por dez conjuntos, agregados por nível de competência e similaridade nas atividades executadas, descritos na Tabela 1. A CBO 2002 toma como base, com algumas adaptações, a metodologia da classificação e agregação internacional de informações ocupacionais de 1988 a *International Standard Classification of Occupations* (ISCO 88) que utiliza a escolaridade como critério para determinar o nível de competência. Os graduados cuja ocupação na CBO tenha nível de escolaridade exigido menor que o ensino superior são considerados sobreeducados e, assim, a *dummy* de *overeducation* será igual a um.

As variáveis que utilizamos no modelo (2) têm como fonte a RAIS. O salário corresponde ao rendimento médio anual do trabalhador, deflacionado pelo Índice Nacional de Preços ao consumidor. A experiência corresponde ao tempo de emprego

²O MTE define o sistema de classificação da CBO 2002 como um documento normalizador do reconhecimento para fins classificatórios, sem função de regulamentação profissional, da nomeação e da codificação dos títulos e conteúdos das ocupações do mercado de trabalho brasileiro para fins estatísticos de registros administrativos, censos populacionais e outras pesquisas domiciliares.

Tabela 1. Grande Grupo CBO 2002.

Código	Título
0	Membros das Forças Armadas, Policiais e Bombeiros Militares
1	Membros Superiores do Poder Público, Dirigentes de Organizações de Interesse Público e de Empresas, Gerentes
2	Profissionais das Ciências e das Artes
3	Técnicos de Nível Médio
4	Trabalhadores de Serviços Administrativos
5	Trabalhadores dos Serviços, Vendedores do Comércio em Lojas e Mercados
6	Trabalhadores Agropecuários, Florestais e da Pesca
7	Trabalhadores da Produção de Bens e Serviços Industriais
8	Trabalhadores da Produção de Bens e Serviços Industriais
9	Trabalhadores em Serviços de Reparação e Manutenção

Fonte: Elaboração própria a partir da Classificação Brasileira de Ocupações.

do trabalhador em meses. Embora tenhamos selecionados indivíduos graduados, introduzimos aqui a variável anos de estudo, pois após a conclusão da graduação alguns indivíduos aumentam seu grau de instrução, fazendo especializações, mestrado e/ou doutorado.

A **Tabela 2** mostra um perfil sócioeconômico dos trabalhadores corretamente alocados (*matched*) e os trabalhadores *overeducated*. São apresentadas médias das variáveis utilizadas nas estimações, bem como a composição das duas amostras no tocante ao gênero, à raça, à idade e características das firmas.

Os trabalhadores *overeducated* apresentam uma média de salário inferior aos *matched*. Embora seja uma mera comparação de médias, mostra um indício da penalização salarial gerada pelo excesso de qualificação. Ambos os grupos apresentam médias de idade e de duração do curso universitário bastante próximas. Os *overeducated* possuem mais tempo de emprego em média e a maioria é composta por indivíduos do sexo feminino, não brancos e oriundos de escolas públicas. Esses grupos, em especial os não brancos de escolas públicas, apresentam historicamente piores condições de renda e educação, de maneira que esses fatores tendem a contribuir para que obtenham menores ganhos salariais e ao mesmo tempo para sua situação de vulnerabilidade social, os conduzindo a ocupações com qualificação requerida inferior.

A **Tabela 3** mostra a média das variáveis incluídas no modelo empírico que estimamos para cada janela de tempo. Acompanhando a evolução das médias salariais, percebe-se uma elevação, embora a diferença existente entre os corretamente alocados e os *overeducated* tenha se mantido. A desigualdade salarial aumentou de uma janela a outra e os egressos com excesso de qualificação tiveram mais tempo de experiência entre as janelas e em comparação com o grupo *matched*. Destaque-se que muitos egressos estão presentes em mais de uma janela, pela interseção entre os anos.

Tabela 2. Perfil Sócioeconômico – Corretamente Alocados (*Matched*) e *Overeducated*.

		<i>Matched</i>	<i>Overeducated</i>
		Média	
Salário		1.211,38	953,32
Idade		28,08	27,4
Anos de Estudo		16,03	16
Tempo de Emprego		33,8	35,8
Variáveis Sócioeconômicas		%	
Sexo	Feminino	48,3	50,2
	Masculino	51,7	49,8
Cor	Branco	52,6	45,8
	Não Brancos	47,4	54,2
Rede Escolar Nível Médio	Particular	51,4	42,2
	Pública	48,6	57,8
Renda Familiar	Menos de R\$ 622,00	31,8	34,5
	De R\$ 622,00 a R\$ 1.865,99	39,0	39,6
	R\$ 1.866,00 ou mais	29,2	25,8
Escolaridade do Pai	Analfabeto	26,88	34,2
	Fundamental	13,28	17,1
	Médio	31,51	28,3
	Superior ou mais	28,33	20,3
CBO 2002 Ano de Formação	Desempregado	69,9	64,4
	CBO1	2,9	0,8
	CBO2	4,5	0,6
	CBO3	3,9	5,4
	CBO4	10,0	14,9
	CBO5	5,1	7,7
	CBO7	1,2	2,0
	CBO8	1,0	1,8
	CBO9	1,5	2,5
Emprego Temporário	Não	99,7	99,5
	Sim	0,3	0,5
CNAE 2.0	Agropecuária e Pesca	0,1	0,1
	Indústria	8,4	12,5
	Construção Civil	2,2	2,4
	Comércio	11,8	14,9
	Serviços	77,4	70,1
Porte da Empresa	Micro	17,8	21,4
	Pequena	15,2	16,2
	Média	14,4	16,1
	Grande	52,7	46,3
Primeiro Emprego	Não	81,0	85,5
	Sim	19,0	14,4
Nascido em João Pessoa	Não	50,9	46,8
	Sim	49,1	53,2

Fonte: Elaboração própria a partir dos dados da STI/UFPB e da RAIS.

Tabela 3. Média das Variáveis por Janela.

	Janelas					
	<i>t</i> = 1	<i>t</i> = 2	<i>t</i> = 3	<i>t</i> = 4	<i>t</i> = 5	<i>t</i> = 6
<i>Matched</i>						
Salário	858,9	918,24	1.014,7	1.114	1.176,11	1.270,91
Anos de Estudo	16	16,5	16,08	17,01	16,02	16,04
Tempo de Emprego	39,4	34,09	30,46	29,66	29,71	29,17
<i>Overeducated</i>						
Salário	703,52	745,6	806,9	868,3	954,16	1057,35
Anos de Estudo	16	17,001	16,04	16	16,06	16
Tempo de Emprego	32,3	32,75	33,72	35,20	36,51	37,1

Fonte: Elaboração própria a partir dos dados da STI/UFPB e da RAIS.

5. Resultados

Esta seção apresenta os impactos do *overeducation* sobre os ganhos salariais dos graduados na UFPB, encontrados com os procedimentos descritos anteriormente. Na análise descritiva dos dados feita na seção anterior, os trabalhadores *overeducated* apresentaram um salário médio menor que seus pares corretamente alocados. Aqui adicionamos técnicas para o controle do viés de seleção afim de fornecer maior robustez aos indícios apontados pelas diferenças de média salarial.

A *Tabela 4* traz os valores dos coeficientes estimados via mínimos quadrados considerando quatro especificações. Na primeira, estimamos a equação minceriana modificada com um *pooled OLS*, sem considerar os efeitos fixos do empregado, da firma e das interações de ambos com o tempo. De um modelo ao outro cada um desses elementos foram incluídos até chegar na especificação mais completa do modelo 4. A variável dependente é o logaritmo do salário médio recebido pelo empregado e *overeducation* é a variável de interesse, com valor 1 para o indivíduo *overeducated* em determinado ano dentro da janela.

Considerando todas as especificações o coeficiente do *overeducation* foi negativo e significativo. Ou seja, os trabalhadores de nível superior em ocupações que exigem menor escolaridade auferem menor salário quando comparados aos trabalhadores com a mesma qualificação em postos de trabalho coerentes com sua formação. Com a introdução dos efeitos fixos, nota-se uma redução na magnitude dos coeficientes, indicando que um procedimento de identificação que não considera a influência de fatores não-observados, como no caso do *pooled OLS* tende a viesar o valor das estimativas.

Quando os efeitos das interações com o tempo são considerados, o viés gerado pelo *OLS* simples fica ainda mais evidente. Nessa especificação, indivíduos *overeducated* têm uma redução de 0,1183 unidades monetárias no seu salário médio quando comparados aos trabalhadores corretamente alocados. Os resultados encontrados são consonantes com a maior parte da literatura, cujas pesquisas mostram uma redução salarial associada ao *overeducation*. Analisando os resultados de [Bauer \(2002\)](#) e [Chevalier \(2003\)](#), vemos algumas diferenças, embora ambos também lancem mão de métodos para tratar a heterogeneidade não observada. [Chevalier \(2003\)](#) também mostra um

Tabela 4. Retorno Salarial do *overeducation* – Estimação por OLS.

Variáveis	Variável Dependente: Log do salário médio			
	Modelos			
	(1)	(2)	(3)	(4)
<i>Overeducation</i>	-0,1831 *** (0,0201)	-0,1261 *** (0,0209)	-0,1435 *** (0,0228)	-0,1183 *** (0,0225)
Anos de Estudo	0,2977 *** (0,0412)	0,1027 *** (0,0368)	0,0217 (0,0328)	0,0158 (0,0458)
Experiência	0,0055 *** (0,0003)	0,0007 * (0,0004)	0,0036 *** (0,0003)	0,001 ** (0,0005)
Experiência ²	-0,0001 ** (0,0003)	-0,00004 (0,00001)	-0,00006 *** (0,00001)	-0,00001 *** (0,00002)
Janelas (<i>Dummies</i>)	Sim	Sim	Sim	Sim
Empregado (FE)	Não	Sim	Sim	Sim
Firma (FE)	Não	Não	Sim	Sim
Interações	Não	Não	Não	Sim
N.obs.	(4.598)	(4.598)	(4.598)	(4.598)
R ²	0,2304	0,8313	0,9279	0,9432

Nota: Elaboração própria a partir das estimações.

Fonte: Erro-padrão entre parênteses. * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$.

efeito negativo do excesso de qualificação, porém com uma magnitude maior. Já Bauer (2002), que utilizou um painel com efeitos fixos do indivíduo, não encontrou diferença estatisticamente significativa entre os salários dos *overeducated* e seus pares.

A experiência gera ganhos de salários até um ponto a partir do qual os rendimentos do trabalhador começam a diminuir. Já os anos de estudo não apresentaram efeito estatisticamente significativo na especificação mais completa. Muitos egressos analisados permaneceram apenas com o curso de graduação, embora alguns tenham aumentado seus anos de estudo com cursos de pós-graduação.

Na Tabela 5 tem-se as estimações por área de formação dos egressos. Os cursos de graduação da UFPB são agrupados em centros de acadêmicos de acordo com as similaridades das áreas de estudo dos cursos. O Centro de Ciências Médicas e da Saúde, por exemplo, reúne os cursos de medicina, biomedicina, nutrição e etc., enquanto que o Centro de Ciências Sociais Aplicadas (CCSA) contém os cursos ligados a negócios, como Administração, Ciências Contábeis e Economia. Sendo assim, tal qual em Annegues et al. (2018), consideramos os centros acadêmicos como as áreas de formação dos egressos. Na área de Humanas consideramos os egressos dos cursos do Centro de Ciências Humanas Letras e Artes (CCHLA), do Centro de Educação (CE) e do Centro de Ciências Jurídicas. Nas exatas, os egressos do Centro de Ciências Exatas e da Natureza (CCEN) e do Centro de Tecnologia (CT), o qual reúne os cursos de Engenharia. Na área da Saúde, os egressos dos Centros de Ciências Médicas e da Saúde (CCM e CCS).

Tabela 5. Retorno Salarial do *overeducation* por área de formação.

Variáveis	Variável Dependente: Log do salário médio			
	Áreas de Estudo			
	Humanas	Exatas	Saúde	Sociais Aplicadas
<i>Overeducation</i>	-0,0203 (0,0430)	-0,2010*** (0,0463)	-0,1487** (0,068)	-0,0177 (0,0472)
Anos de Estudo	0,0256** (0,011)	0,0361 (0,0577)	0,1113 (0,1262)	0,1809* (0,0911)
Experiência	0,004** (0,0011)	0,0025 (0,0004)	-0,00024 (0,002)	0,0064*** (0,0013)
Experiência ²	-0,0001*** (0,00005)	-0,00004 (0,00001)	-0,00001 (0,00001)	-0,00002*** (0,00004)
Janelas (<i>Dummies</i>)	Sim	Sim	Sim	Sim
Empregado (FE)	Sim	Sim	Sim	Sim
Firma (FE)	Sim	Sim	Sim	Sim
Interações	Sim	Sim	Sim	Sim
N.obs.	761	605	714	647
R ²	0,9415	0,9295	0,9522	0,9677

Nota: Elaboração própria a partir das estimações.

Fonte: Erro-padrão entre parênteses. * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$.

Apenas nas áreas de Exatas e Saúde foram identificadas diferenças significativas de salário em razão do *overeducation*. O alto retorno associado a estas profissões e a exigência de certificações para o seu exercício, as quais definem bem as ocupações condizentes com a área de formação, podem explicar este resultado.

Além dos resultados por área de formação, também estimamos os resultados por gênero, raça, rede de ensino no qual o egresso cursou o ensino médio e escolaridade dos pais (Tabela 6). Em todas as categorias houve redução estatisticamente significativa de salário. As mulheres apresentaram maior magnitude de coeficiente, sendo mais penalizadas em razão da condição de *overeducated*.

Com relação à raça, brancos e não brancos apresentaram praticamente os mesmos coeficientes. No tocante ao *Background Familiar*, aqui medido pela escolaridade do pai, os *overeducated* filhos de pais com ensino superior tiveram uma redução no salário de 0,2075, resultado próximo dos filhos de pais com ensino fundamental. Os *overeducated* que estudaram em escolas particulares tiveram maior coeficiente relativamente aos oriundos de escolas públicas.

Mesmo com os controles pelos efeitos fixos do indivíduo e da firma e das interações desses efeitos com o tempo, queremos testar o papel das habilidades individuais na determinação dos salários. Para isso, seguindo a estratégia sugerida por McGuinness (2006), estimamos o impacto do *overeducation* por quantis. Supondo que os quantis de salário refletem os níveis de habilidades, ao comparar a diferença de salário entre *overeducated* e corretamente alocados dentro de cada quantil, comparamos indivíduos com

Tabela 6. Resultado OLS por gênero, raça, rede de ensino e escolaridade dos pais.

Variáveis	<i>Overeducation</i>	Erro-Padrão	R ²
Homens	-0,0899***	0,0323	0,9516
Mulheres	-0,1297***	0,0320	0,9330
Branco	-0,1108***	0,0305	0,9536
Não Branco	-0,1140***	0,0331	0,9348
Escolaridade do Pai			
Analfabeto	-0,0250	0,0382	0,9456
Ensino fundamental	-0,1996***	0,0542	0,9640
Ensino médio	-0,1200***	0,0453	0,8852
Ensino superior ou mais	-0,2075***	0,0448	0,9292
Rede onde cursou o ensino médio			
Pública	-0,1212***	0,0298	0,9366
Privada	-0,1334***	0,0348	0,9509

Nota: Elaboração própria a partir das estimações.

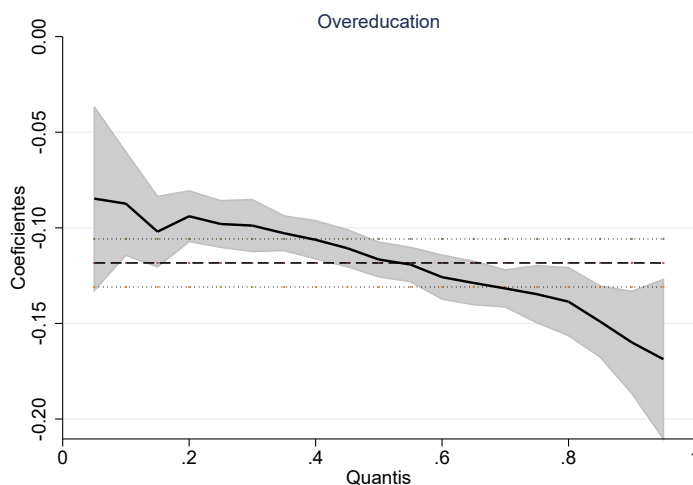
Fonte: Erro-padrão entre parênteses. * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$.

mesma habilidade. Além disso, a estimação na média pode mascarar a variabilidade de efeitos do *overeducation* entre os diversos pontos da distribuição salarial.

O gráfico da [Figura 1](#) mostra os coeficientes estimados ao longo dos quantis de salário. São plotadas no gráfico as estimativas da dummy de *overeducation* para cada quantil entre 0,05 e 0,95, bem como o intervalo de confiança de 95%, representado pela área cinza em volta da curva. Já a linha horizontal representa o efeito na média condicional dos salários. Tal qual as estimativas por mínimos quadrados, o *overeducation* apresentou efeito negativo em todos os quantis. Analisando a magnitude dos efeitos, vemos que os indivíduos localizados nos menores quantis têm a menor diferença de salário em relação aos corretamente alocados.

Ao longo da distribuição o efeito se torna cada vez mais negativo, de maneira que os indivíduos nos maiores quantis de salário, que seriam os mais habilitados de acordo com nossa hipótese levantada acima, são os que apresentaram maior redução de salários em razão do excesso de educação. Ou seja, existem indícios que vão contra a tese de que as diferenças de rendimento entre corretamente alocados e *overeducated* estão mais ligadas às diferenças de habilidades não observadas. O artigo de [McGuinness \(2006\)](#), por outro lado, mostra evidências em favor dessa hipótese, uma vez que os coeficientes se mostraram mais negativos para aqueles localizados nos quantis inferiores.

Mesmo estando em ocupações que exigem menos escolaridade, a maior habilidade dos egressos nos quantis superiores poderia conferir um ganho salarial extra em relação aos demais trabalhadores da função e compensar a diferença de rendimento médio relativamente aos indivíduos de nível superior em postos de trabalhos condizentes com a escolaridade. No entanto, os resultados por quantis sugerem que essa compensação não ocorre.



Fonte: Elaboração própria a partir das estimações.

Figura 1. Retorno Salarial do *overeducation* – Regressão Quantílica.

6. Considerações Finais

O presente artigo investigou como o *overeducation*, que é o excesso de qualificação em relação ao que é requerido pelo emprego, afeta o salário recebido. Dada a recente expansão da oferta de trabalhadores com nível superior ocorrida no Brasil, analisamos o retorno salarial do *overeducation* para indivíduos graduados na UFPB. Identificamos estes egressos na Relação Anual de Informações Sociais (RAIS), acompanhando-os entre os anos de 2003 e 2013.

O ponto central na identificação do efeito do *overeducation* sobre o salário é o possível viés de seleção dos indivíduos *overeducated*. Fatores não observados, em especial as habilidades inatas, podem determinar tanto o salário do egresso quando a situação de estar em uma ocupação que não requer o seu grau de instrução. Nesse sentido, a estratégia de identificação que adotamos consiste em uma estimação com efeitos fixos do indivíduo e da firma na qual trabalha e interações destes fatores com o tempo. Essas interações surgem como *proxies* das características não observadas variantes no tempo, uma vez que métodos mais sofisticados, como o uso de variáveis instrumentais, nem sempre são factíveis em razão da disponibilidade de dados. Para testar a hipótese de que a desigualdade de salário entre *overeducated* e corretamente alocados é produto das diferenças de habilidade, a equação de salários foi estimada via regressão quantílica. Supondo que a posição relativa dos indivíduos na distribuição de salários identifique o seu nível de habilidade, se o *overeducation* penaliza mais os egressos de quantis inferiores, há indícios de que a habilidade exerce um papel importante na determinação dos salários vis-à-vis a condição de *overeducation*.

Em suma, os resultados mostraram que egressos *overeducated* têm uma redução salarial quando comparados com seus pares cuja ocupação corresponde à sua escolaridade. Tal redução se mostrou mais forte para indivíduos formados nas áreas de ciências exatas e saúde, e do sexo feminino. Já os resultados da estimação por quantis mostrou que o *overeducation* afeta negativamente mais os salários dos indivíduos localizados no topo da distribuição de salários, ou seja, aqueles com maior nível de habilidade.

Desse modo, há indícios de que a habilidade parece não ser determinante para o menor salário auferido pelos *overeducated*.

As habilidades inatas destes trabalhadores parecem não ser suficientes a ponto de lhes fornecer um prêmio salarial próximo ao que teriam caso estivessem em uma ocupação que exige o ensino superior. As evidências que encontramos seguem a mesma direção de boa parte desta literatura e mostram que o *overeducation* consiste em um sinal de ineficiência dos gastos em educação, uma vez que estes nem sempre se traduzem em ganhos de renda e maior mobilidade social.

Referências bibliográficas

- Annegues, A. C., Oliveira, C., Figueiredo, E., & Porto, S., Jr. (2018, 4–5 de julho). Overeducation e área de formação: Evidências para os egressos da UFPB. In *XXIII Encontro Regional de Economia*, Fortaleza, CE. https://www.anpec.org.br/nordeste/2018/submissao/arquivos_identificados/057-30545e4aac19d96be6b75873119c721d.pdf
- Baltagi, B. H., Egger, P. H., & Pfaffermayr, M. (2014, janeiro). *Panel data gravity models of international trade* (CESifo Working Paper N° 4616). Munich, Germany: CESifo. https://www.ifo.de/DocDL/cesifo1_wp4616.pdf
- Bauer, T. K. (2002). Educational mismatch and wages: A panel analysis. *Economics of Education Review*, 21(3), 221–229. [http://dx.doi.org/10.1016/S0272-7757\(01\)00004-8](http://dx.doi.org/10.1016/S0272-7757(01)00004-8)
- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, Part 2), 9–49. <https://www.jstor.org/stable/1829103>
- Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3), 368–386. <http://dx.doi.org/10.1111/j.1368-423X.2011.00349.x>
- Carneiro, A., Guimarães, P., & Portugal, P. (2009, maio). *Real wages and the business cycle: Accounting for worker and firm heterogeneity* (IZA Discussion Paper N° 4174). Bonn, Germany: Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA). <http://ftp.iza.org/dp4174.pdf>
- Chevalier, A. (2003). Measuring over-education. *Economica*, 70(279), 509–531. <http://dx.doi.org/10.1111/1468-0335.t01-1-00296>
- Cohn, E., & Ng, Y. C. (2000). Incidence and wage effects of overschooling and underschooling in Hong Kong. *Economics of Education Review*, 19(2), 159–168. [http://dx.doi.org/10.1016/S0272-7757\(99\)00006-0](http://dx.doi.org/10.1016/S0272-7757(99)00006-0)
- Diaz, M. D. M., & Machado, L. (2008). Overeducation e undereducation no Brasil: Incidência e retornos. *Estudos Econômicos*, 38(3), 431–460. <http://dx.doi.org/10.1590/S0101-41612008000300001>
- Dolton, P. J., & Silles, M. A. (2008). The effects of over-education on earnings in the graduate labour market. *Economics of Education Review*, 27(2), 125–139. <http://dx.doi.org/10.1016/j.econedurev.2006.08.008>
- Dolton, P. J., & Vignoles, A. (2000). The incidence and effects of overeducation in the U.K. graduate labour market. *Economics of Education Review*, 19(2), 179–198. [http://dx.doi.org/10.1016/S0272-7757\(97\)00036-8](http://dx.doi.org/10.1016/S0272-7757(97)00036-8)
- Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*, 1(1), 75–86. [http://dx.doi.org/10.1016/0272-7757\(81\)90028-5](http://dx.doi.org/10.1016/0272-7757(81)90028-5)
- Freeman, R. (1976). *The overeducated American*. Academic Press.

- Groot, W. (1996). The incidence of, and returns to overeducation in the UK. *Applied Economics*, 28(10), 1345–1350. <http://dx.doi.org/10.1080/000368496327895>
- Guimarães, P., & Portugal, P. (2009, janeiro). *A simple feasible alternative procedure to estimate models with high-dimensional fixed effects* (IZA Discussion Paper N° 3935). Bonn, Germany: Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA). <http://ftp.iza.org/dp3935.pdf>
- Kiersztyn, A. (2013). Stuck in a mismatch? The persistence of overeducation during twenty years of the post-communist transition in Poland. *Economics of Education Review*, 32, 78–91. <http://dx.doi.org/10.1016/j.econedurev.2012.09.009>
- Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. In *Handbook of the economics of education* (Vol. 4, pp. 283–326). Amsterdam: Elsevier. <http://dx.doi.org/10.1016/B978-0-444-53444-6.00003-1>
- McGuinness, S. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, 20(3), 387–418. <http://dx.doi.org/10.1111/j.0950-0804.2006.00284.x>
- Mincer, J. (1974). *Schooling, experience, and earnings* [Human Behavior & Social Institutions, Vol. 2]. New York, NY: National Bureau of Economic Research. <https://www.nber.org/books/minc74-1>
- Morano, C. P. (2014). The determinants of overeducation : Evidence from the Italian labour market Preliminary Draft. *Investigaciones de economía de la educación*, 9, 681–698. <http://repec.economicsofeducation.com/2014valencia/valencia2014.pdf>
- Oliveira, C., Mariano, F. Z., & Araújo, J. A. (2019, 18–19 de julho). A qualidade e expansão do ensino superior contribui para o overeducation? Evidências para o Ceará. In *XXIV Encontro Regional de Economia*, Fortaleza, CE. https://www.anpec.org.br/nordeste/2019/submissao/arquivos_identificados/054-b43a5baf3e680a1d9e70b2e1ca5157eb.pdf
- Reimer, D., Noelke, C., & Kucel, A. (2008). Labor market effects of field of study in comparative perspective: An analysis of 22 European countries. *International Journal of Comparative Sociology*, 49(4-5), 233–256. <http://dx.doi.org/10.1177/0020715208093076>
- Rumberger, R. W. (1981). The rising incidence of overeducation in the U.S. labor market. *Economics of Education Review*, 1(3), 293–314. [http://dx.doi.org/10.1016/0272-7757\(81\)90001-7](http://dx.doi.org/10.1016/0272-7757(81)90001-7)
- Santos, A. M. d. (2002). Overeducation no mercado de trabalho brasileiro. *Revista Brasileira de Economia de Empresas*, 2(2), 1–22. <https://portalrevistas.ucb.br/index.php/rbee/article/view/4377>
- Sattinger, M. (1993). Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31(2), 831–880. <https://www.jstor.org/stable/2728516>
- Schank, T., Schnabel, C., & Wagner, J. (2010). Higher wages in exporting firms: Self-selection, export effect, or both? First evidence from linked employer-employee data. *Review of World Economics*, 146(2), 303–322. <http://dx.doi.org/10.1007/s10290-010-0049-7>
- Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, 51(1), 1–17. <https://www.jstor.org/stable/1818907>
- Sicherman, N. (1991). “Overeducation” in the labor market. *Journal of Labor Economics*, 9(2), 101–122. <http://dx.doi.org/10.1086/298261>
- Smith, J. P., & Welch, F. (1978, novembro). *The overeducated American? A review article* (The Rand Paper Series N° RAND-P-6253; ERIC ED175340). Santa Monica, CA: Rand Corp. <https://eric.ed.gov/?id=ED175340>
- Thurow, L. C. (1975). *Generating inequality: Mechanisms of distribution in the US economy*. Basic Books.

Verdugo, R. R., & Verdugo, N. T. (1989). The impact of surplus schooling on earnings: Some additional findings. *Journal of Human Resources*, 24(4), 629–643.
<http://dx.doi.org/10.2307/145998>