

***Overeducation* e Área de Formação: evidências para os Egressos da UFPB**

Ana Cláudia Annegues,^a * Celina Oliveira^{c†} Erik Figueiredo^{b‡} Sabino Porto Júnior,^a §

^a Universidade Federal do Rio Grande do Sul ^b Universidade Federal da Paraíba ^c Universidade Federal do Ceará

Área 13 - Economia do Trabalho

Resumo

O objetivo do presente artigo consiste em investigar o efeito da área de formação dos egressos da UFPB sobre a probabilidade deste estar em uma ocupação com nível de instrução exigido abaixo da sua formação superior, ou seja, numa situação de *overeducation*. Investigou-se também a influência das áreas sobre o tempo que esses indivíduos permanecem nesta condição. Para isso, utilizou-se um logit com *Propensity Score Matching* para múltiplos tratamentos e modelos de análise de sobrevivência. Os resultados mostraram que os formados nas áreas de humanidades, artes e sociais aplicadas apresentam maior probabilidade de estarem sob *overeducation*, bem como uma menor chance de deixar essa condição durante alguns anos após a conclusão do curso.

Palavras-chave: *Overeducation*; Áreas de Estudo

Classificação JEL: I23, C51

Abstract

The aim of the present paper is to investigate the effect of training area of UFPB graduates on the probability of being in an occupation with required level of education below their higher education, i.e. in an *overeducation* situation. We also investigated the influence of areas on the time these individuals remain in this condition. For this, we used a logit with *Propensity Score Matching* for multiple treatments and survival analysis models. The results showed that graduates in humanities, arts and applied social sciences are more likely to be *overeducated*, as well as a lower chance of leaving this condition for a few years after completion of the course.

Keywords: *Overeducation*; Areas of Study

JEL Codes: I23, C51

*annegues.ana@gmail.com.

†oli.celina@gmail.com.

‡eafigueiredo@gmail.com.

§sabino@ppge.ufrgs.br.

1 Introdução

Durante as últimas décadas verificou-se uma forte expansão na oferta de trabalhadores com ensino superior no Brasil. Segundo dados do Censo, o percentual de brasileiros com diploma universitário aumentou de 4,4% em 2000 para 7,9% em 2010. Embora haja um alto retorno salarial associado à formação em nível superior [ver Sachsida, Loureiro e Mendonça (2004); Resende e Wyllie (2006); Pessôa e Filho (2008); Teixeira e Menezes-Filho (2012)], nem sempre a posse do diploma universitário representa uma ferramenta de ascensão social. Isto porque existem indícios de que o mercado de trabalho não tem conseguido absorver o número crescente de trabalhadores qualificados, levando estes a ocupações que exigem um grau de instrução menor que o possuído. Este fenômeno é denominado pela literatura de sobreeducação ou *overeducation*.

Segundo Groot (1996), o excesso de qualificação pode gerar consequências adversas para a sociedade em termos de subutilização do capital humano e ineficiência dos gastos em educação. Para os trabalhadores *overeducated*, é possível que estes obtenham rendimentos inferiores aos obtidos por trabalhadores que desempenham atividades condizentes com seu nível de formação, desmotivando-o a melhorar sua capacitação, tendo em vista que o investimento em qualificação não possui o retorno esperado na forma de salário. Nesse sentido, o *overeducation* tem sido associado a uma menor produtividade e a baixa satisfação do trabalho [ver Groot e Brink (2000)].

Não obstante, as causas e as consequências do *overeducation* ainda não são totalmente claras, dado que não há uma teoria unânime sobre o tema. Até o momento, a literatura tem interpretado o fenômeno utilizando algumas teorias sobre 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*.

Por originar-se de um desequilíbrio entre oferta e demanda no mercado de trabalho, a consistência do *overeducation* com a teoria do capital humano baseia-se na hipótese desse fenômeno como sendo de curto prazo, onde o mercado naturalmente se encarregaria de corrigir os desajustes entre firmas e trabalhadores. Já sob a perspectiva dos modelos de competição por emprego, a rigidez da demanda por trabalhadores qualificados levaria estes a acumularem mais anos de estudo como forma de assegurar sua posição na competição pela vaga de trabalho, em alguns casos mais do que o necessário para uma determinada ocupação. Nessas condições, o *overeducation* tende a se tornar uma condição permanente aos indivíduos.

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).

Assim, a tese do *overeducation* como sendo de fato um problema pode depender de um aspecto fundamental, que é a duração do fenômeno. Os trabalhadores podem ocupar temporariamente funções que exigem menos da sua formação por uma necessidade imediata ou pela expectativa de ascender de função no futuro. No caso de indivíduos recém-formados, é possível que muitos estudantes iniciem o curso universitário já trabalhando em empregos de nível médio e permaneçam nessas ocupações mesmo após a conclusão da graduação, enquanto buscam um *match* adequado. Para Groot (1996) os trabalhadores *overeducated* são relativamente improdutivos, quando comparados com os corretamente alocados entre instrução e ocupação. Dolton e Vignoles (2000) encontraram evidências de que muitos graduados em situação de *overeducation* no primeiro emprego permaneceram nessa mesma ocupação seis anos após o término do

curso.

Alguns estudos apontam que a permanência no *overeducation* pode estar ligada também a natureza de algumas áreas de formação [ver Dolton e Vignoles (2000) e Reimer, Noelke e Kucel (2008)]. O mercado de trabalho de áreas como artes, filosofia e sociologia, por exemplo, tenderia a ser mais restrito, de modo que muitos graduados não encontram espaço para atuar nesses campos de estudo. Consequências semelhantes podem ocorrer quando há a criação de novos cursos universitários em que não se leva em conta as possibilidades de atuação nesses novos profissionais no mercado de trabalho.

Dito isso, o objetivo do presente estudo consiste em fornecer evidências empíricas sobre a influência dos campos de estudo na probabilidade de um indivíduo ser *overeducated* e na duração dessa situação após a conclusão do curso. A introdução da dimensão de tempo na análise procura captar os fatores que explicam a permanência de alguns graduados em ocupações que não exigem o ensino superior. Não sendo este um problema transitório, torna-se imperioso conhecer a dimensão dessa influência, uma vez que a criação e manutenção dos cursos universitários, bem como outras políticas de expansão do ensino superior, implicam em custos à sociedade.

A estratégia empírica a ser adotada baseia-se em dois procedimentos. A escolha da área de atuação por parte dos indivíduos não é aleatória, de modo que fatores omitidos podem estar relacionados ao seu estado de *overeducated*. Assim, para mensurar a probabilidade de ocorrer *overeducation*, estimou-se um modelo logit com pareamento via *Propensity Score Matching*, afim de reduzir a influência dos chamados *confounders*. Para a duração do *overeducation*, empregou-se a abordagem da Análise de Sobrevivência, tendo em vista a natureza dos dados de duração. Esta técnica busca mensurar a chance de um graduado *overeducated* sair dessa condição, dada a sua área de formação e demais variáveis.

Esses procedimentos foram aplicados aos 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 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 campos de estudo são representados por *dummies* para cada um dos centros acadêmicos da UFPB, os quais contém cursos de áreas profissionais semelhantes. São analisados também o papel de variáveis acadêmicas (desempenho, duração do curso), atributos individuais (sexo, raça) e características da ocupação.

O artigo encontra-se estruturado da seguinte forma. Na seção 2 são apresentadas algumas evidências presentes na literatura empírica. A seção 3 traz os detalhes dos procedimentos metodológicos adotados. A seção 4 mostra os dados utilizados e uma análise descritiva. A seção 5 traz os resultados encontrados. Por fim, a seção 6 contém as considerações finais.

2 Literatura Empírica

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. 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. Como resultado encontraram que a educação requerida possui um retorno salarial duas vezes maior que a educação possuída pelo indivíduo.

Dentre os determinantes micro e macroeconômicos do *overeducation* comumente analisados pela liter-

atura 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 entendimento da literatura de que a qualidade do ajuste entre instrução e ocupação aumenta com a idade do indivíduo.

Sobre a influência do gênero no *overeducation*, argumenta-se que as mulheres têm uma maior probabilidade de estarem sobreeducadas do que os homens, pois as alternativas de trabalho disponíveis para as mulheres são mais restritas, como os trabalhos em meio expediente (OFEK; MERRILL, 1997). Porém, McGoldrick e Robst (1996) e Morano (2014) encontram evidências empíricas de que não há diferença de gênero na probabilidade do trabalhador se encontrar em *overeducation*.

Algumas variáveis macroeconômicas podem desempenhar um papel importante sobre a situação de *overeducation*, no sentido de que condições macroeconômicas adversas tendem a restringir as oportunidades de emprego ao trabalhador, fazendo com que ele aceite ocupar um posto, mesmo que este não corresponda ao seu nível de formação (DOLTON; SILLES, 2008). Liu, Salvanes e Sørensen (2016) mostram que um aumento de 1 ponto percentual na taxa de desemprego aumenta em 3,4% a probabilidade de estar sobreeducado (MORANO, 2014).

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; KUCCEL, 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. Barone e Ortiz (2011) investigam o papel de características da formação superior sobre a probabilidade de *overeducation* entre os recém-formados de universidades europeias e mostram que graduados das áreas de humanas e dos cursos de bacharelado estão mais expostos ao *overeducation*.

3 Procedimentos Metodológicos

3.1 Modelos Empíricos

O objetivo do presente artigo consiste em mensurar a influência do campo de estudo dos graduados na UFPB sobre a probabilidade destes se encontrarem numa situação de *overeducation* e sobre o tempo de duração destes egressos em ocupações cuja escolaridade requerida está abaixo da sua formação superior. Para tanto, serão estimados dois modelos empíricos. O primeiro consiste em um logit cuja variável dependente é uma *dummy* para o status do egresso i em um determinado ano t (OE_{it}); 1 se *overeducated* e 0 caso contrário:

$$P(OE_{it} = 1 | Centros_i, X_i) = G(\beta_0 + \beta_1 X_i + Centros_i) \quad (1)$$

O campo de estudo é representado pelo centro acadêmico no qual o indivíduo estudou. Foram incluídas *dummies* para cada um dos centros da UFPB, denotadas na equação (1) por $Centros_i$. São eles: o Centro de

Ciências Humanas, Letras e Artes (CCHLA), o Centro de Comunicação, Turismo e Artes (CCTA), o Centro de Ciências Sociais Aplicadas (CCSA), o Centro de Ciências Jurídicas (CCJ), o Centro de Educação (CE), o Centro de Ciências Exatas e da Natureza, o Centro de Informática (CI), o Centro de Tecnologia (CT) e o Centro de Ciências da Saúde (CCS)¹.

Os cursos que compõem cada centro estão no Anexo I. O CCHLA corresponderia ao campo de estudo das humanidades (Filosofia, Sociologia, etc.). O CCSA representa áreas ligadas a negócios, com cursos como Economia e Administração, e o CCJ representa a área do Direito. Já o CCEN representa o campo das ciências exatas (Matemática, Estatística e etc.), o CI a área de Tecnologia e Informática, o CT as engenharias e o CCS representa o campo da Saúde.

X_i é um vetor que inclui variáveis comumente presentes na literatura, como características individuais (sexo, raça) e da ocupação (tempo empregado, se o emprego é temporário) e variáveis macroeconômicas (taxa de desemprego). Além dessas, foram analisadas algumas variáveis acadêmicas, como o Coeficiente de Rendimento Escolar (CRE), que mostra o desempenho acadêmico do aluno na universidade, e a duração do curso universitário.

No segundo modelo, estimamos a influência desses mesmos determinantes sobre o tempo de duração do indivíduo no estado de *overeducation* como variável dependente. Neste caso, utilizou-se um modelo de duração ou Análise de Sobrevivência. Através do modelo de duração é possível estimar a probabilidade de um indivíduo sair do *overeducation* e encontrar uma ocupação a qual exija o nível de instrução possuído pelo trabalhador. As subseções 3.2 e 3.3 explanam com mais detalhes os métodos de estimação de cada um dos dois modelos.

3.2 *Propensity Score Matching* para múltiplos tratamentos

A principal hipótese a ser testada aqui é se os campos de estudo dos graduados na UFPB impactam a probabilidade desses indivíduos se encontrarem na situação de *overeducation*. Conforme já explicitado na introdução, a distribuição dos alunos entre os cursos da universidade não se dá de forma aleatória, mas sim pelas escolhas feitas pelos mesmos, as quais são determinadas por fatores que podem simultaneamente estar relacionados ao fato de serem sobreeducados. Uma forma de eliminar o viés de seleção seria encontrar alguma variável instrumental que não estivesse diretamente correlacionada com a variável de resultado. No entanto, encontrar um instrumento exógeno nem sempre é uma tarefa trivial.

Outra estratégia para identificar efeitos causais é o pareamento via *Propensity Score Matching* (PSM). A principal hipótese deste método postula que ao controlar por um vetor de variáveis observáveis os resultados potenciais se tornam independentes da variável de tratamento (hipótese de seleção nas observáveis ou ignorabilidade). Em outras palavras, comparar grupos igualmente elegíveis a receber determinado tratamento, dado um conjunto de características observáveis, reduz ou elimina o viés de seleção (ROSENBAUM; RUBIN, 1983).

Seja Y_i a variável de resultado observada para o indivíduo i e sejam T_i a variável de tratamento e Y^1 e Y^0 os resultados potenciais dos grupos de tratamento e controle, respectivamente. Tem-se, então:

$$Y_i = T_i Y_i^1 + (1 - T_i) Y_i^0 \quad (2)$$

Pela hipótese de seleção nas observáveis, $Y_i^1, Y_i^0 \perp T_i | X_i$ implica que $Y_i^1, Y_i^0 \perp T_i | p(X_i)$, em que $p(X_i)$ é o escore de propensão, que consiste na probabilidade de participação no tratamento condicionada ao vetor de características observáveis X_i . A estimação do escore de propensão é obtida através de algum modelo de

¹Foi incluído no CCS o Centro de Ciências Médicas (CCM), que contém o curso de medicina

escolha qualitativa, como um logit. Desse modo, o efeito do tratamento sobre os tratados (ATT) pode ser obtido por:

$$\beta_{ATT} = E(Y_i^1 - Y_i^0) \quad (3)$$

$$\beta_{ATT} = E\{E[Y_i|p(X_i), T_i = 1] - E[Y_i|p(X_i), T_i = 0]\} \quad (4)$$

Uma forma de encontrar o efeito do tratamento é considerar o escore de propensão como elemento ponderador na regressão da variável de resultado contra T_i . A ponderação é dada por:

$$w(T, X) = T_i + (1 - T_i) \frac{\hat{p}(X_i)}{1 - \hat{p}(X_i)} \quad (5)$$

Esta abordagem é utilizada quando se tem apenas um status de tratamento que permite separar os indivíduos entre duas categorias: tratados e não tratados. No caso do presente artigo, temos mais de uma modalidade de tratamento, que são cada um dos centros acadêmicos nos quais os indivíduos se graduaram na UFPB. Assim, é necessário adaptar o arcabouço de resultados potenciais, bem como a estimação dos escores de propensão no PSM, para mais categorias de tratamento.

Seja M o número de tratamentos cujos efeitos se quer mensurar. Nesse estudo, $M = 9$ corresponde ao número de centros acadêmicos da UFPB. Os resultados potenciais de um indivíduo são representados por Y_i^T , onde $T = 1, \dots, M$. O efeito causal de cada tratamento é definido como a diferença entre os resultados potenciais de cada um para um mesmo indivíduo. Isto é, $D(T', T'') = Y_i^{T'} - Y_i^{T''}$ para todo $T' \neq T''$.

Alguns papers teóricos mostram que é possível estender o uso do PSM para mais de uma condição de interesse, como Imbens (2000) e Imai e Dyk (2004). Spreeuwenberg et al. (2010) sugere a utilização de um logit ou um probit multinomial para estimar os escores de propensão. Não obstante, conforme apontado por McCaffrey et al (2013), a implementação prática das abordagens para múltiplos tratamentos ainda é escassa na literatura. Nesse sentido, os autores fornecem um guia para implementação do PSM para múltiplos tratamentos propondo a utilização de uma técnica de *marching learning* para estimação dos pesos com os escores de propensão, o *Generalized Boosted Model*(GBM).

A estimação por GBM envolve um processo iterativo com árvores de regressões múltiplas para capturar relações complexas e não lineares entre o atributo de tratamento e o vetor de características observáveis (McCaffrey et al, 2013). O algoritmo inicia com uma regressão simples e a cada nova interação uma árvore com outra regressão é adicionada e cada interação melhora o grau de ajustamento do modelo. Para evitar o problema de *overfitting*, o GBM seleciona um número razoável de interações afim de atender determinados critérios. No PSM, o critério consistiria no balanceamento do vetor de variáveis observáveis entre os grupos de tratamento e controle. Em outras palavras, utiliza-se o GBM para estimar os escores de propensão, com um número ótimo de interações que minimize a diferença entre a distribuição dos escores de ambos os grupos.

Em termos de redução de viés, alguns estudos sugerem que o GBM apresenta melhores resultados comparativamente aos modelos de regressão logística [ver Lee, Lessler e Stuart (2010)]. Outras vantagens do GBM, segundo McCaffrey et al (2013) são: i) permite a inclusão de um maior número de variáveis observáveis no modelo; ii) funciona para variáveis de tratamento discretas e contínuas e iii) é invariante a transformações monotônicas nas variáveis.

A extensão do uso do GBM para múltiplos tratamentos sugerida por McCaffrey et al (2013) baseia-se nos seguintes passos: i) primeiro criam-se variáveis *dummy* para cada uma das M categorias de tratamento; ii) em seguida, ajusta-se um GBM para cada um dos indicadores binários de tratamento que foram criados,

para obter a probabilidade de submeter-se aquele tratamento, dado o vetor de características observáveis e iii) por fim calcula-se os pesos a serem utilizados para obter o efeito dos tratamentos. Intuitivamente, a estimação dos pesos requer o conhecimento da probabilidade de receber um determinado tratamento com relação aos demais. No contexto do presente artigo, estimamos o modelo logit da equação (1) com a reponderação pelos escores de propensão estimados via GBM.

3.3 Análise de Sobrevivência

A análise de duração do *overeducation* após a conclusão do curso superior requer técnicas especiais devido a estrutura dos dados. No presente artigo, a variável de interesse é o tempo decorrido até a ocorrência de um evento em particular, que é a saída do egresso da situação de *overeducation*, o que caracteriza um caso de dados de sobrevivência.

Sendo assim, assumir uma distribuição normal para esses tipos de dados não seria uma escolha razoável, uma vez que estes geralmente não são distribuídos simetricamente, aparecendo positivamente distorcidos e com poucas observações sobrevivendo durante muito tempo em comparação com a maioria. Ademais, algumas observações podem não ter atingido o ponto final de interesse - deixar a situação de *overeducation*-, de modo que o tempo de sobrevivência exato não será conhecido, caracterizando uma censura à direita.

A partir disso, a análise de dados do tempo de sobrevivência baseia-se em duas funções utilizadas para descrever sua distribuição: a função de sobrevivência e a função de risco.

Função de Sobrevivência: como colocado por Everitt e Hothorn (2009), a função de sobrevivência $S(t)$ é definida como a probabilidade de que o tempo de sobrevivência, T , seja maior ou igual a algum tempo t , ou seja, $S(t) = P(T \geq t)$. Na presença de dados censurados, a função de sobrevivência não paramétrica é estimada usando o estimador de Kaplan-Meier [Kaplan e Meier (1958)].

O primeiro passo do estimador de Kaplan-Meier envolve a ordenação dos tempos de sobrevivência do menor para o maior, de modo que $t_1 \leq t_2 \leq \dots \leq t_n$, onde o último elemento t_j representa o maior tempo de sobrevivência. A partir disso, a estimativa Kaplan-Meier da função de sobrevivência é obtida como segue:

$$(\hat{S})(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{r_j}\right) \quad (6)$$

onde d_j é o número de indivíduos que experimentam o evento de interesse no tempo t_j e r_j é o número de indivíduos em risco antes de t_j , incluindo aqueles censurados. Como exemplo, a função de sobrevivência para o segundo tempo que ocorre o evento de interesse, t_2 é igual à probabilidade estimada de não morrer no tempo t_2 , condicionada a fato do indivíduo estar ainda em risco no tempo t_2 .

Uma das vantagens do método não paramétrico, como colocado por Cameron e Trivedi (2005), é a não necessidade de especificar qualquer função de risco e considerar que os tempos de sobrevivência observados são independentes e identicamente distribuídos, permitindo ainda um diagnóstico rápido do formato da função de sobrevivência.

Função de Risco: a função de risco (*Hazard function*) $h(t)$ é definida como a probabilidade de um indivíduo experimentar o evento em um pequeno intervalo de tempo, s , dado que o indivíduo tenha sobrevivido até o início do intervalo. Esta função é dada por:

$$h(t) = \lim_{s \rightarrow 0} P(t \leq T \leq t + s | T \geq t) \quad (7)$$

onde T é o tempo de sobrevivência do indivíduo.

Com a função de risco é possível verificar quais fatores afetam o risco de ocorrência do evento, ou seja, a saída da situação de *overeducation*. Desse modo, condiciona-se a função *hazard* a um conjunto de covariáveis x que representam as características dos indivíduos, da firma na qual estão empregados e das condições de mercado de trabalho.

Como já mencionado, a variável resposta é o tempo de sobrevivência na situação de *overeducation*, que é possivelmente censurada. Devido a isso, faz-se o uso de uma técnica de regressão específica para modelar a relação da variável resposta com as variáveis explicativas, chamada de Regressão Cox, introduzida por Cox (1972), que faz uso da função de risco para a modelagem. Como $h(t)$ é estritamente positivo, adequa-se para uma função linear das variáveis explicativas da seguinte forma:

$$\log(h(t)) = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q \quad (8)$$

Segundo Cox (1972), a função 8 é adequada apenas para uma função de risco constante ao longo do tempo, o que na prática é pouco provável de ocorrer. Este problema foi superado com o modelo de risco proporcional proposto também por Cox (1972), permitindo que a forma de dependência de $h(t)$ sobre t permaneça não especificada, de modo que:

$$\log(h(t)) = \log(h_0(t)) + \beta_1 x_1 + \dots + \beta_q x_q \quad (9)$$

onde $h_0(t)$ é uma função de risco *baseline* com todas as variáveis explicativas iguais a zero. Pode-se reescrever o modelo como:

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_q x_q) \quad (10)$$

No modelo Cox, a função de risco *baseline* descreve a forma comum da distribuição do tempo de sobrevivência para todos os indivíduos, enquanto a função de risco relativo dada por $\exp(\beta_1 x_1 + \dots + \beta_q x_q)$, dá o nível de risco de cada indivíduo.

Os parâmetros em um modelo de Cox podem ser estimados maximizando a probabilidade parcial derivada ao assumir os tempos de sobrevivência contínuos. (KALBFLEISCH; PRENTICE, 1980)

A especificação final para estimação do modelo de duração do *mismatch* é dada por:

$$\text{duração}_{\text{overeducation}} = X\beta \quad (11)$$

onde x é o conjunto de variáveis explicativas como, por exemplo, sexo, cor, rede escolar, renda familiar, grau de instrução do pai, ocupação, se o emprego é temporário, o tempo no emprego, se está no primeiro emprego, se nasceu em João Pessoa, duração do curso, CRE final do curso, taxa de desemprego, idade, remuneração média, área dos cursos de formação de nível superior dada pelo centro acadêmico, atividade econômica e porte da firma.

4 Dados

O banco de dados aqui utilizado é resultado da junção de informações acadêmicas dos egressos fornecidas pela Superintendência de Tecnologia da Informação (STI) da UFPB e pela Comissão Permanente de Vestibular (COPERVE), 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.

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 informações socioeconômicas coletadas no

momento do vestibular pela COPERVE. 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.

No presente estudo, empregou-se 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 de classificação da Classificação Brasileira de Ocupações (CBO) de 2002. Segundo 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 trabalho².

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 no quadro 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.

Quadro 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

A nossa amostra é composta por indivíduos que concluíram sua graduação na UFPB no período de 2003 a 2013. Foram selecionados os concluintes de cada ano em toda UFPB e os mesmos foram acompanhados

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.

desde o ano de conclusão do curso até o último ano analisado. Assim, 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 4.612 observações.

A tabela 1 traz estatísticas que descrevem a amostra considerada na presente pesquisa, indivíduos com nível superior pela Universidade Federal da Paraíba acompanhados em um painel a partir do ano de 2003 até 2013. Considerando tanto a amostra completa como os indivíduos apenas na situação de *Overeducation* incluindo todos os anos do painel, observa-se que os indivíduos são relativamente jovens com média de idade de 27 anos.

Quanto ao Coeficiente de Rendimento Escolar (CRE)³, Duração do Curso e Taxa de Desemprego, a média se manteve a mesma tanto para a amostra completa quanto para os indivíduos na situação de *Overeducation*, apresentando as seguintes médias: 7,9 para o CRE, 4,9 e 5 ano para a duração de curso e 7,8 e 7,9 para a Taxa de Desemprego. Com relação a média salarial e Tempo de Emprego em uma firma, na amostra completa apresentou uma média salarial maior e menor tempo de vínculo empregatício em relação aqueles indivíduos que permaneceram na situação de *Overeducation*.

Observando as variáveis categóricas, a amostra completa apresenta em sua maioria mais homens, indivíduos que se consideram não brancos, que estudaram em rede pública de nível médio com renda familiar na faixa de R\$ 622,00 a R\$ 1.865,99. A amostra completa apresentou também indivíduos com pai analfabeto e formandos pela UFPB na situação de desempregados. Além disso, boa parte das observações dos que estão empregados estão em vínculos não temporários, no setor de atividade de Serviços em firmas de grande porte, com 19% da amostra caracterizado como primeiro emprego e 50,9% nascidos fora do município de João Pessoa.

Tabela 1: Estatística Descritiva para a Amostra Completa e para os *Overeducated*

		Completa	Overeducation
Variáveis Contínuas		Média	
	Idade	27,7	27,4
	Salário	1061,08	953,32
	CRE	7,9	7,9
	Duração do Curso	4,9	5,0
	Tempo de Emprego	33,8	35,8
	Taxa de Desemprego	7,8	7,9
Variáveis Categóricas		%	
Sexo	Feminino	49,4	50,2
	Masculino	50,6	49,8
Cor	Branco	48,7	45,8
	Não Branco	51,3	54,2
Rede Escolar Nível Médio	Particular	46,0	42,2
	Pública	54,0	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
	De R\$ 1.866,00 ou mais	29,2	25,8

³O CRE corresponde à média de notas nas disciplinas que o aluno obteve durante o curso

Escolaridade do Pai	Analfabeto	31,2	34,2
	Fundamental	15,5	17,1
	Médio	29,7	28,3
	Superior ou mais	23,7	20,3
CBO 2002	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

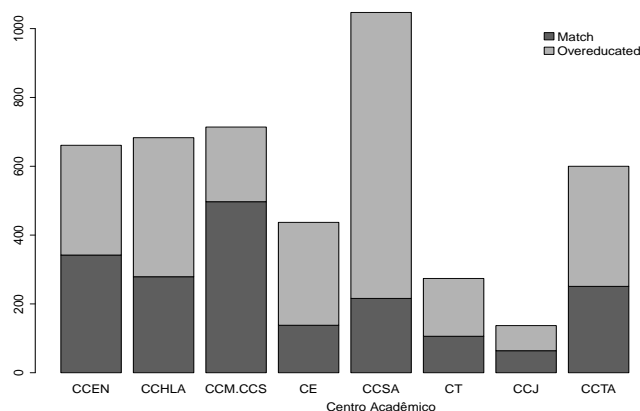
Por fim, as variáveis categóricas para os indivíduos na situação de *Overeducation* mostram que 50,2% são mulheres e 54,2% se declaram não brancos. Além disso, a grande maioria desses indivíduos cursou o ensino médio em escolas públicas, com renda familiar na faixa de R\$ 622,00 a R\$ 1.865,99 e pai analfabeto.

Destaca-se ainda que grande parte dos indivíduos estavam desempregados na situação de formados pela UFPB. Daqueles que estavam empregados, apenas 0,5% eram emprego temporário e grande parte dos indivíduos em *Overeducation* estão no setor de Serviços e em empresas de grande porte. Além disso, 14,4% dos indivíduos em *Overeducation* estão em seu primeiro emprego e 53,2% nasceram em João Pessoa.

A figura 1 contém um gráfico com a distribuição dos egressos em cada centro. A parte cinza das barras representa os indivíduos em situação de *overeducation*, enquanto que a porção mais escura representa aqueles que possuem um *match* adequado entre ocupação e escolaridade.

O CCSA, que consiste na área de negócios e economia, é o que apresenta maior proporção de ex-alunos em situação de *overeducation* no mercado de trabalho. Outros centros cuja maior parte dos egressos está em empregos que não exigem nível superior são o CCTA, ou a área de artes e comunicação, o CCHLA, da área de humanas, o Centro de Educação, CE, e o CT, que inclui os cursos de Engenharia.

Figura 1: Proporção de egressos *overeducated* por Centro Acadêmico



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

No CCEN (Exatas) e no Centro de Informática, a distribuição dos ex-alunos entre as duas categorias se mostra mais equilibrada. Na área de saúde e medicina, representada pela coluna CCM.CCS, apresenta a menor parcela de egressos *overeducated*

A tabela 2 mostra que um egresso da UFPB passa, em média, mais de 3 anos em empregos não condizentes com sua formação. Comparando-se a média de duração entre os centros acadêmicos, os ex-alunos formados nos Centros de Educação e de Ciências Sociais Aplicadas são os que passam mais tempo na situação de *overeducation*, com um tempo médio de mais de 4 e 5 anos, respectivamente. Os egressos Centro de Ciências Jurídicas e dos Centros de Saúde, são os que passam menos tempo pelo fenômeno.

Tabela 2: Duração média dos egressos em *overeducation*

Centros Acadêmicos	Duração Média
CCEN	2,98
CCHLA	3,81
CCM.CCS	1,55
CE	4,18
CCSA	5,16
CT	2,77
CCJ	2,19
CCTA	3,39
CI	3,33
UFPB	3,51

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

5 Resultados

Esta seção apresenta as estimações do efeito dos campos de estudo sobre a situação de *overeducation* e sobre a duração do fenômeno após a conclusão do curso de graduação. A subseção 5.1 mostra o resultado da estimação do modelo logit com PSM para múltiplos tratamentos e a seção 5.2 traz os resultados encontrados pela abordagem de análise de sobrevivência.

5.1 Logit com PSM para múltiplos tratamentos

O primeiro passo do PSM consistiu em estimar via GBM as probabilidades de pertencer aos centros acadêmicos condicionadas a um vetor de características observáveis pré-tratamento. Foram incluídas nesse vetor as seguintes variáveis: sexo, raça, rede escolar em que cursou o ensino médio, renda familiar, escolaridade dos pais, se possui computador e acesso à internet e estado civil.

Após o pareamento via escore de propensão, é relevante testar o balanceamento das características observáveis entre os egressos dos centros acadêmicos. Para tanto, o algoritmo de McCaffrey (2013) fornece como medida de balanceamento a diferença das médias das variáveis em cada grupo, comparando um centro acadêmico com o outro, dois dois.

Tendo em vista o grande número de linhas em razão do número de modalidades de tratamento e de características observáveis, não inserimos a tabela com as diferenças de média⁴ Em todas as variáveis houve uma redução importante na diferença entre os grupos antes e depois do pareamento. Na variável sexo, por exemplo, a condição de interesse é ser do sexo masculino e a diferença entre a proporção de homens no CCHLA e no CCSA reduziu de 0,78 para 0,33 após o pareamento.

As categorias de tratamento são os centros da UFPB nos quais os egressos se formaram. Defiu-se o CCEN como centro base sobre o qual se compara a probabilidade de um egresso dos demais centros ser *overeducated* (categoria omitida). A tabela 3 traz as estimações dos determinantes do *overeducation*. A primeira coluna mostra as estimações para o logit simples, sem o pareamento. No tocante aos centros acadêmicos, quase todos tiveram efeito significativo sobre a probabilidade de ser *overeducated*, exceto o CCJ, o qual representa o campo de estudo do direito. Os formados das Ciências Humanas, Sociais Aplicadas e Artes e Comunicação têm maior probabilidade de estarem na situação de *overeducation*, em comparação com os egressos das Ciências Exatas (categoria base). O mesmo ocorre com os formados na área de educação e nas engenharias (CT). Já os estudantes das áreas de Saúde e Informática possuem um menor risco de estarem em ocupações cujo grau de instrução requerido seja inferior à sua formação universitária.

Tabela 3: Determinantes do *overeducation*

Variáveis	LOGIT	LOGIT-PSM
CCHLA	0.4409*** (0.1414)	0.3728*** (0.0560)
CCM_CCS	-0.8218*** (0.1433)	-0.9768*** (0.0574)
CE	0.5740*** (0.1702)	0.5224*** (0.0666)
CCSA	1.1277*** (0.1330)	1.0374*** (0.0590)
CT	0.5440*** (0.1850)	0.6887*** (0.0620)
CCJ	0.5759** (0.2353)	-0.0803 (0.0763)
CCTA	0.4137*** (0.1437)	0.3677*** (0.0566)
CI	-0.4198 (0.3601)	-0.4699*** (0.0772)

⁴A tabela encontra-se disponível mediante solicitação aos autores.

Homem	-0.2644*** (0.0806)	-0.4425*** (0.0308)
Não Brancos	-0.0239 (0.0751)	0.0299 (0.0295)
Nível Médio Escola Pública	0.0209 (0.0854)	-0.0490 (0.0332)
De R\$ 622,00 a R\$ 1.865,99	-0.0203 (0.0922)	-0.1596*** (0.0366)
De R\$ 1.866,00 ou mais	-0.0561 (0.1111)	-0.1279*** (0.0432)
Fundamental	-0.2388** (0.1213)	-0.1695*** (0.0497)
Médio	-0.3995*** (0.1016)	-0.4855*** (0.0404)
Superior ou Mais	-0.5082*** (0.1162)	-0.3538*** (0.0453)
Emprego Temporário	3.8233** (1.5335)	4.7545*** (0.8130)
Pequena Empresa	-0.1610 (0.1366)	-0.2090*** (0.0567)
Média Empresa	0.0320 (0.1444)	-0.0314 (0.0602)
Grande Empresa	-0.1187 (0.1218)	-0.1481*** (0.0520)
Primeiro Emprego	-0.8166*** (0.0932)	-0.8352*** (0.0360)
Nascido em João Pessoa	0.3664*** (0.0760)	0.1922*** (0.0302)
Duração do Curso	0.0761** (0.0333)	0.0342*** (0.0130)
CRE	-0.0121 (0.0523)	-0.1143*** (0.0212)
Salário	-0.0001*** (0.00004)	-0.0001*** (0.00002)
Tempo de Vínculo	0.0048*** (0.0012)	0.0048*** (0.0005)
Taxa de Desemprego	0.3119*** (0.0380)	0.2512*** (0.0152)
Idade	-0.0342*** (0.0086)	-0.0447*** (0.0035)
C	-1.6690 (1.1084)	0.8306** (0.4015)
Dummies CBO 2002	Sim	Sim
Dummies CNAE 2.0	Sim	Sim
Observações	4612	4612
Log Likelihood	-2302.22	-15060.05
Akaike Inf. Crit.	4686.44	30202.11

Fonte: elaboração própria com base nas estimações.

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Os resultados para os indivíduos das áreas de ciências humanas e artes estão em consonância com as evidências existentes, já discutidas na seção 2. O mercado de trabalho mais restrito característico desses campos de atuação pode gerar um excesso de oferta de trabalhadores, de modo que muitos acabam sendo alocados para funções não condizentes com sua instrução. O resultado para educação e engenharia, mesmo com um mercado de trabalho mais amplo, pode ser sintomático do excesso de trabalhadores qualificados gerados pela expansão do ensino superior nos últimos anos. O excedente de trabalhadores não consegue ser absorvido pelo mercado, gerando uma subutilização desses indivíduos.

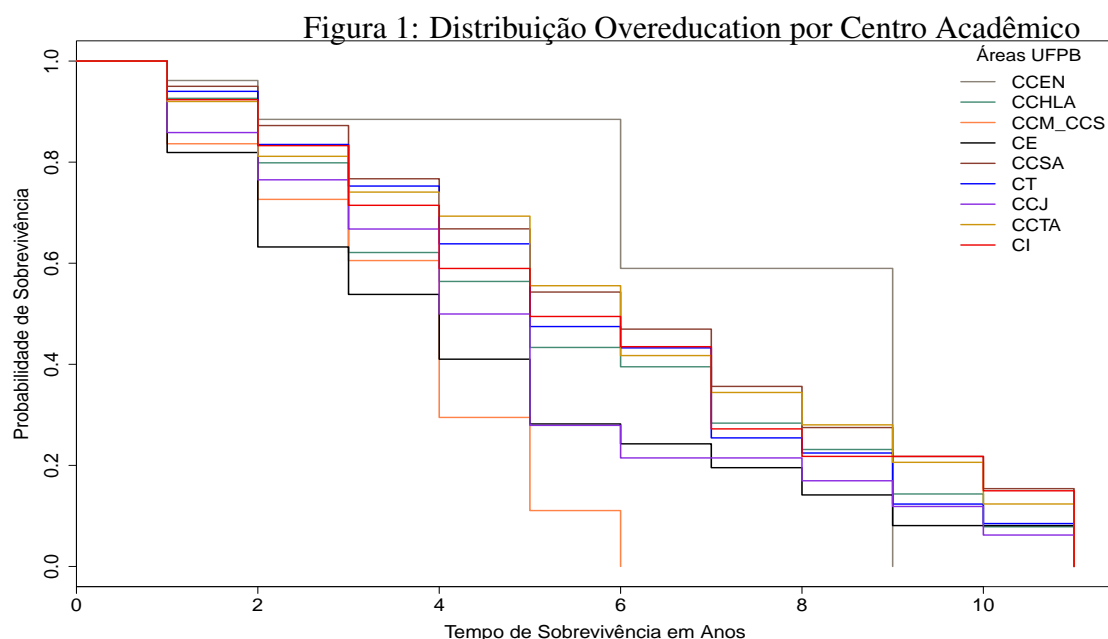
No tocante aos demais determinantes, alguns deles já foram objeto de análise da literatura sobre o tema. As estimativas mostram que: as mulheres e os indivíduos mais jovens têm maior probabilidade de serem *overeducated*. Conforme já mostrado na seção 2, esse é um resultado consensual na literatura, a qual argumenta que os trabalhadores com estas características têm um espaço de oportunidades mais reduzido em comparação com os demais. Mesmo com as mudanças culturais ocorridas nas últimas décadas, grande parte da mulheres ainda acumulam junto à sua atividade profissional as responsabilidades da casa e da criação dos filhos, de modo que tendem a aceitar ocupações que lhe permitam se dedicar a essas funções, mesmo que exijam menos da sua formação.

A taxa de desemprego têm uma influência positiva sobre essa probabilidade, na medida em que ela reduz as alternativas de trabalho disponíveis aos egressos, que acabam aceitando empregos com nível de instrução exigido menor.

Com relação às variáveis acadêmicas, as evidências sugerem que quanto mais baixo o desempenho acadêmico (CRE) e maior o tempo de duração do curso, maiores as chances de estar em *overeducation*.

5.2 Análise da Duração do *Overeducation*

A análise de duração para a situação de *Overeducation* se divide em duas abordagens, a paramétrica e a não paramétrica. No modelo não paramétrico, a partir do estimador de Kaplan-Meier é possível estimar a função de sobrevivência exposta na figura 2 por centro acadêmico.



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

Pode-se ver que até o 9º ano os indivíduos formados no CCEN possuem uma maior probabilidade de continuar na situação de *Overeducation*, seguido dos formados no CCSA e no CCTA. Já os centros CCM.CCS e CCJ, os centros de ciências médicas e jurídicas, respectivamente, apresentaram baixas probabilidades de continuar na situação de *Overeducation*, em síntese, indivíduos formados nessas áreas possuem maiores chances de se ocuparem em cargos que exijam nível superior.

Ressalta-se que este resultado faz parte de uma análise preliminar e não informa sobre quais fatores levam à saída do estado de *Overeducation*. Em seguida, a tabela 4 complementa a análise e expõe quais os determinantes da duração do *Overeducation* a partir do modelo de regressão Cox.

As estimativas mostraram que atributos como sexo, cor, idade, rede escolar de nível médio, renda familiar, pai com grau de instrução analfabeto, a ocupação no ano conclusão do curso, e alguns centros como o CCHLA, CCM.CCS, CT e CCJ são os principais condicionantes da duração de *Overeducation* por apresentarem sinais estatisticamente significantes. Os resultados informam que se o indivíduo for homem, não branco, e se cursou o ensino médio na rede pública têm menos chances de sair da situação de *overeducation*.

Outras características como ter pai analfabeto, renda familiar de R\$ 622,00 a R\$ 1.865,99 e trabalhar em uma firma de porte médio também contribuem negativamente para a probabilidade de saída da situação de *Overeducation*. O salário, o tempo de vínculo na firma, a taxa de desemprego e a idade também contribuem de forma negativa para a saída desse status. Já fatores como emprego temporário e estar no primeiro emprego contribuem de forma positiva para a saída da condição de *Overeducated*.

Em suma, os indivíduos menos privilegiados em termos socioeconômicos são mais suscetíveis a realizar um *match* inadequado. Este resultado chama a atenção, pois embora o foco das políticas de expansão do ensino superior seja a qualificação dessas pessoas com vistas a proporcionar maior mobilidade social, os egressos terminam ocupando empregos não compatíveis com a formação universitária.

Analisando os centros acadêmicos, as variáveis chave da presente pesquisa, verifica-se que ser formando nos centros CCM.CCS, CT e CCJ contribui de forma positiva para a saída da situação de *Overeducation*. Já os formados no CCHLA têm menos chances de deixar o *overeducation* e conseguir outra ocupação que forneça um bom *match* com sua escolaridade.

Tabela 4: Resultados para Duração do *overeducation*

Variáveis	Coefficientes	Desvio Padrão
CCHLA	-0.2091**	(0.0845)
CCM.CCS	0.4711***	(0.0981)
CE	0.0728	(0.0915)
CCSA	-0.0811	(0.0712)
CT	0.5490***	(0.1067)
CCJ	0.6756***	(0.1427)
CCTA	-0.0197	(0.0863)
CI	-0.3898	(0.2510)
Homem	-0.1809***	(0.0452)
Não Brancos	-0.2639***	(0.0436)
Nível Médio Escola Pública1	-0.2009***	(0.0496)
De R\$ 622,00 a R\$ 1.865,99	-0.1902***	(0.0502)
De R\$ 1.866,00 ou mais	-0.1785***	(0.0675)
Fundamental	-0.1247**	(0.0626)
Médio	0.0202	(0.0560)
Superior ou Mais	-0.0701	(0.0713)
Pequena Empresa	-0.0065	(0.0670)
Média Empresa	-0.2023***	(0.0690)
Grande Empresa	0.0474	(0.0620)
Primeiro Emprego	0.2496***	(0.0592)
Nascido em João Pessoa	0.0484	(0.0438)
Duração do Curso	-0.0285	(0.0174)
CRE	-0.0007	(0.0275)
Salário	-0.0001***	(0.00003)
Tempo de Vínculo	-0.0077***	(0.0007)
Taxa de Desemprego	-0.2222***	(0.0187)
Idade	-0.0306***	(0.0049)
Dummies CBO 2002	Sim	Sim
Dummies CNAE 2.0	Sim	Sim
R ²		0.3286
Score (Logrank) Test	1572.3060***	(df = 40)

Fonte: elaboração própria com base nas estimações.

Nota: *p<0.1; **p<0.05; ***p<0.01

6 Considerações Finais

Um trabalhador encontra-se em *overeducation* quando o seu grau de instrução é superior ao exigido pelo seu emprego. Esses indivíduos tendem a auferir rendimentos menores que aqueles que desempenham funções compatíveis com sua formação, embora possuam a mesma escolaridade. Há consequências adversas também para a sociedade, na forma de subutilização dos seus recursos. O objetivo desta pesquisa consistiu em investigar os determinantes desse fenômeno e do seu tempo de duração entre trabalhadores qualificados egressos da UFPB. Em particular, buscou-se mensurar o efeito das áreas de formação (Humanidades, Direito, Economia), representadas pelos centros acadêmicos da universidade. Para isso, estimou-se um logit com *Propensity Score Matching* para mensurar a probabilidade de *overeducation*. E para o modelo de duração utilizou-se a abordagem de análise

de sobrevivência.

Os resultados mostraram que ser das áreas de Humanidades, Artes, Sociais Aplicadas, Educação e Engenharia aumenta a probabilidade de um egresso da UFPB ser um trabalhador *overeducated*. Já análise de duração mostrou que a probabilidade de deixar a situação de *overeducation* após um certo tempo se reduz para os formados nessas áreas, exceto para os formados em Engenharia. Os egressos das áreas de Saúde e Direito apresentaram menor probabilidade de serem *overeducated* e aqueles que o são possuem menor chance de permanecer nessa condição.

Vale ressaltar que os resultados encontrados constituem evidências para os egressos da UFPB, o que pode comprometer a validade externa da pesquisa, isto é, o grau de aplicabilidade ou generalização do estudo a outros contextos. Nesse sentido, trabalhar com dados de egressos do Brasil consistiria em um avanço importante para futuras pesquisas, tendo em vista as diferenças econômicas entre os mercados de trabalho das regiões.

Outro avanço seria mensurar o custo do *overeducation* para os recursos públicos. O diploma universitário de um aluno implica em custos para a sociedade, como o pagamento de professores e manutenção de estrutura física, portanto o *overeducation* sinaliza um ponto de ineficiência na alocação desses recursos. As políticas educacionais voltadas à expansão do ensino superior precisam ser complementadas com ações que melhor direcionem o crescente número de trabalhadores qualificados às demandas do mercado de trabalho.

Referências

BARONE, C.; ORTIZ, L. Overeducation among european university graduates: a comparative analysis of its incidence and the importance of higher education differentiation. *Higher Education*, Springer, v. 61, n. 3, p. 325–337, 2011.

BECKER, G. S. Investment in human capital: A theoretical analysis. *Journal of political economy*, The University of Chicago Press, v. 70, n. 5, Part 2, p. 9–49, 1962.

CAMERON, A. C.; TRIVEDI, P. *Microeconometrics*. Cambridge University Press, 2005. Disponível em: <<https://EconPapers.repec.org/RePEc:cup:cbooks:9780521848053>>.

COX, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, Royal Statistical Society, v. 34, n. 2, p. 187–220, 1972.

DOLTON, P.; VIGNOLES, A. The incidence and effects of overeducation in the uk graduate labour market. *Economics of education review*, Elsevier, v. 19, n. 2, p. 179–198, 2000.

DOLTON, P. J.; SILLES, M. A. The effects of over-education on earnings in the graduate labour market. *Economics of Education Review*, Elsevier, v. 27, n. 2, p. 125–139, 2008.

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

EVERITT, B. S.; HOTHORN, T. *A Handbook of Statistical Analyses Using R, Second Edition*. 2nd. ed. [S.l.]: Chapman & Hall/CRC, 2009.

FREEMAN, R. *The Overeducated American*. [S.l.: s.n.], 1976.

GROOT, W. The incidence of, and returns to overeducation in the uk. *Applied Economics*, Taylor & Francis, v. 28, n. 10, p. 1345–1350, 1996.

GROOT, W.; BRINK, H. M. V. D. Overeducation in the labor market: a meta-analysis. *Economics of education review*, Elsevier, v. 19, n. 2, p. 149–158, 2000.

IMAI, K.; DYK, D. A. V. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, Taylor & Francis, v. 99, n. 467, p. 854–866, 2004.

IMBENS, G. W. The role of the propensity score in estimating dose-response functions. *Biometrika*, Oxford University Press, v. 87, n. 3, p. 706–710, 2000.

KALBFLEISCH, J. D.; PRENTICE, R. L. *The Statistical Analysis of Failure Time Data (Wiley Series in Probability and Statistics)*. 1st. ed. Wiley-Interscience, 1980. Hardcover. ISBN 0471055190. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0471055190>>.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, American Statistical Association, Taylor & Francis, Ltd., v. 53, n. 282, p. 457–481, 1958.

- KIERSZTYN, A. Stuck in a mismatch? the persistence of overeducation during twenty years of the post-communist transition in poland. *Economics of Education Review*, Elsevier, v. 32, p. 78–91, 2013.
- LEE, B. K.; LESSLER, J.; STUART, E. A. Improving propensity score weighting using machine learning. *Statistics in medicine*, Wiley Online Library, v. 29, n. 3, p. 337–346, 2010.
- LIU, K.; SALVANES, K. G.; SØRENSEN, E. Ø. Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, Elsevier, v. 84, p. 3–17, 2016.
- MCGOLDRICK, K.; ROBST, J. Gender differences in overeducation: A test of the theory of differential overqualification. *The American Economic Review*, JSTOR, v. 86, n. 2, p. 280–284, 1996.
- MCGUINNESS, S. Overeducation in the labour market. *Journal of economic surveys*, Wiley Online Library, v. 20, n. 3, p. 387–418, 2006.
- MORANO, C. P. The determinants of overeducation : Evidence from the Italian labour market Preliminary Draft. n. November, p. 25–27, 2014.
- OFEK, H.; MERRILL, Y. Labor immobility and the formation of gender wage gaps in local markets. *Economic Inquiry*, v. 35, n. 1, p. 28–47, 1997. Disponível em: <<https://EconPapers.repec.org/RePEc:oup:ecinqu:v:35:y:1997:i:1:p:28-47>>.
- PESSÔA, S. de A.; FILHO, F. de H. B. Retorno da educação no brasil. *Pesquisa e Planejamento Econômico*, v. 38, n. 1, 2008.
- REIMER, D.; NOELKE, C.; KUCCEL, A. Labor market effects of field of study in comparative perspective: An analysis of 22 European Countries. *International Journal of Comparative Sociology*, v. 49, n. 4-5, p. 233–256, 2008.
- RESENDE, M.; WYLLIE, R. Retornos para educação no Brasil: evidências empíricas adicionais. *Economia Aplicada*, v. 10, p. 349–365, 2006.
- ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, Oxford University Press, v. 70, n. 1, p. 41–55, 1983.
- SACHSIDA, A.; LOUREIRO, P. R. A.; MENDONÇA, M. J. C. d. Um estudo sobre retorno em escolaridade no brasil. *Revista Brasileira de Economia*, SciELO Brasil, v. 58, n. 2, p. 249–265, 2004.
- SATTINGER, M. Assignment models of the distribution of earnings. *Journal of economic literature*, JSTOR, v. 31, n. 2, p. 831–880, 1993.
- SCHULTZ, T. W. Investment in human capital. *The American economic review*, JSTOR, v. 51, n. 1, p. 1–17, 1961.
- SMITH, J. P.; WELCH, F. The overeducated american? a review article. ERIC, 1978.
- SPREEUWENBERG, M. D. et al. The multiple propensity score as control for bias in the comparison of more than two treatment arms: an introduction from a case study in mental health. *Medical care*, JSTOR, p. 166–174, 2010.
- TEIXEIRA, W. M.; MENEZES-FILHO, N. A. Estimando o retorno à educação do brasil considerando a legislação educacional brasileira como um instrumento. *Revista de Economia Política*, SciELO Brasil, v. 32, n. 3, p. 479–496, 2012.
- THUROW, L. C. *Generating Inequality: Mechanisms of Distribu-Lion in the US Economy*. [S.l.]: Basic Books, 1975.