

# TEXTO PARA DISCUSSÃO N.2 ANO 1



UNIVERSIDADE  
FEDERAL  
DE PERNAMBUCO

## **Universidade Federal de Pernambuco**

Alfredo Macedo Gomes (Reitor)

Moacyr Cunha de Araújo Filho (Vice-Reitor)

## **Programa de Pós-graduação em Economia (PPGECON)**

Roberta de Moraes Rocha (Coordenadora)

Klebson Humberto de Lucena Moura (Vice-Coodenador)

## **Editores da série de Textos para Discussão**

Monaliza Ferreira de Oliveira (PPGECON-UFPE)

<https://www.ufpe.br/ppgecon>

## **Textos para Discussão**

A série de Textos para Discussão divulga resultados preliminares de estudos desenvolvidos na área de Economia.

As opiniões contidas nesta publicação são de exclusiva responsabilidade dos autores, não exprimindo necessariamente o ponto de vista do Programa de Pós-graduação em Economia (PPGECON), da Universidade Federal de Pernambuco.

*Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the publishers and do not necessarily reflect the views of the PPGECON-UFPE.*

## DINÂMICA DO COMPORTAMENTO DISCRIMINATÓRIO NO MERCADO DE TRABALHO BRASILEIRO

Leandro Willer P. Coimbra <sup>1</sup>

Paulo Guilherme Moreira de Melo Filho<sup>2</sup>

Tatiane Almeida de Menezes <sup>3</sup>

### RESUMO

A desigualdade salarial é caracterizada como discriminação quando esta é motivada por crenças ou preferências não ligadas à produtividade individual dos trabalhadores. O objetivo deste artigo é a constatação do comportamento discriminatório no mercado de trabalho relacionado à cor da pele e evidenciar sua evolução temporal. Utiliza-se do método de *Propensity Score Matching* para analisar a diferença média dos ganhos entre "brancos" e "não brancos" através de dados da PNAD de 1996 a 2014. A análise confirma os ganhos médios menores dos trabalhadores "não brancos" ao longo de todo o período. Em 2014 o rendimento se mostrou em torno de 6% menor para os trabalhadores "não brancos". Os dados sugerem ainda que a queda nos diferenciais ao longo dos anos foi estimulada em grande parte pela maior igualdade de oportunidade entre os grupos. Enquanto a discriminação pela cor dentro do mercado de trabalho apresentou queda relativamente menores.

**Palavras-chaves:** Diferença salarial. Discriminação. Mercado de trabalho.

**Classificação JEL:** J7

## 1 Introdução

A literatura referente ao comportamento discriminatório nas ciências econômicas tem início com o trabalho germinal intitulado "Economia da Discriminação". A discriminação é apresentada como um componente da função de preferência dos agentes econômicos, de forma que estes estariam dispostos a pagar uma "taxa" para reduzir o convívio com grupos específicos. Esta é denominada como a teoria da discriminação por preferência (BECKER, 1957)

Outra forma de interpretar a motivação do comportamento discriminatório na visão econômica é através dos modelos estatísticos de discriminação (ARROW, 1973; PHELPS, 1972). Nestes, a discriminação é entendida como resultado de uma estratégia racional sob presença de informação imperfeita. Usa-se de informações livres, como a distribuição de produtividade em grupos de fácil distinção (brancos e pretos, homens e mulheres, deficientes físicos, por exemplo), para condicionar a produtividade esperada de um trabalhador e assim definir seu salário. Neste caso, a diferenciação dos salários é justificada como forma racional do uso das informações disponíveis.

Em suma, a diferenciação entre as duas abordagens está em considerar a discriminação como uma preferência individual ou um mecanismo racional de atuação em um mercado com informação imperfeita. Contudo, em ambas as abordagens, uma consequência da discriminação é que prejudica os incentivos à melhoria da produtividade. Quando os trabalhadores de determinados grupos esperam regras tendenciosas de contratação, isto reflete nos retornos e, conseqüentemente, na decisão de investir em melhorias das habilidades produtivas (COATE; LOURY, 1993). Ou seja, grupos teriam menos incentivos a investir em capital humano, gerando prejuízos sobre o estoque da economia em questão.

<sup>1</sup> Professor Adjunto do departamento de economia da UFPE, Campus do Agreste. Email: leandro.willer@ufpe.br

<sup>2</sup> Professor e pesquisador na Universidade Federal de Pernambuco.

<sup>3</sup> Professora Doutora Associada do Departamento de Economia da UFPE

Na mesma direção em [Lundberg e Startz \(1998\)](#), utiliza-se do conceito de capital social, observado nos modelos de George J. Borjas, para justificar os efeitos alocativos danosos gerados pela discriminação. Na sua visão, cria-se uma espécie de "armadilha da pobreza" para determinados grupos de trabalhadores devido a características produtivas diferentes, não necessariamente reais.

O objetivo deste trabalho é caracterizar a existência do comportamento discriminatório no mercado de trabalho e mensurar a diferença salarial gerada por este comportamento ao longo das últimas décadas no mercado de trabalho brasileiro. Ressalta-se, que diferenças salariais médias entre distintos grupos não se referem necessariamente ao comportamento discriminatório. Caso os trabalhadores mais hábeis se encontrem em maior prevalência em determinado grupo, por exemplo, é natural que a remuneração neste seja maior e justificável do ponto de vista produtivo. Neste caso, o combate a esta diferença poderia gerar um desestímulo ao investimento em melhorias na produtividade semelhante ao impacto discutido da discriminação.

De forma a destacar os impactos da discriminação pela cor, utilizou-se o método de *propensity score matching*. Características de esforço e circunstâncias, como definido em [Roemer \(2002\)](#), estabeleceram grupos de pareamento com trabalhadores com potenciais produtivos semelhantes. A partir da hipótese de que a comparação ocorre entre indivíduos com produtividades (quase) idênticas dentro de cada grupo, obteve-se a razão entre a remuneração de trabalhadores auto declarados de cor preta e parda em relação àqueles autodeclarados de cor branca. A razão salarial menor que um foi então interpretada como um tratamento social diferenciado aos trabalhadores "não brancos" com efeito negativo sobre sua remuneração.

Além desta introdução, o presente trabalho é composto de uma revisão da literatura sobre a persistência do comportamento discriminatório no mercado de trabalho; segue-se da apresentação da estratégia econométrica e dos dados utilizados. A seção 4 discute-se os resultados de forma que esta é dividida em: análise descritiva dos dados, resultado do pareamento, um análise de consistência do resultado. Por fim, são apresentadas as considerações finais.

## 2 Persistência da discriminação no mercado de trabalho

Na abordagem introduzida em [Roemer \(2002\)](#) a discussão da desigualdade salarial é apresentada como resultado das diferenças entre variáveis de esforço e de circunstância. A distinção entre estes conceitos baseia-se na responsabilidade direta, ou não, do trabalhador, por exemplo, enquanto o *background* familiar é uma característica de circunstância, na qual o trabalhador é dotado desta arbitrariamente, o nível de educação ou experiência são, por outro lado, características que representam seu esforço, no qual advém das escolhas deste indivíduo. Vale ressaltar, o efeito de interação no qual resulta em níveis de esforço médio diferentes quando condicionado a diferentes circunstâncias, o que traduz parte do efeito das circunstâncias sobre o diferencial de renda.

A desigualdade salarial passa à análise de discriminação quando a crença do empregador, não necessariamente correta, de que um grupo de trabalhador possui menor produtividade que outro, gera contratações com salários diferenciados em detrimento das reais produtividades individuais dos trabalhadores, como apresentado em [Arrow \(1973\)](#) e [Phelps \(1972\)](#). Nesta visão a discriminação surge de um processo estratégico no qual as características observáveis individuais são complementadas pelas informações do grupo a que se pertence. A análise ocorre sob três hipóteses: 1) fácil ou não custosa distinção entre os grupos de trabalhadores; 2) há um custo para determinação da real produtividade dos trabalhadores, como um período de treinamento (hipótese da assimetria); e 3) uma taxa preconcebida que diferencie a capacidade produtiva dos grupos, como a distribuição de produtividade intragrupo ([ARROW](#),

1973). Portanto, a desigualdade salarial passa a incorporar também fatores como crença e assimetria informacional<sup>4</sup>.

Nestes modelos de discriminação estatística, caso haja correspondência entre as crenças dos empregadores e a realidade do mercado, a discriminação pode ser apontada como comportamento racional e necessário para a eficiência produtiva em caso de assimetria informacional quanto à produtividade dos trabalhadores. Contudo, em [Lundberg e Startz \(1983\)](#) é defendido que a discriminação estatística pode não ser eficiente se ocorre regras de contratações tendenciosas. Isto é, um empregador pode se encontrar em maior proximidade informacional a um grupo que de outro. Neste caso, como no modelo de "investimento em capital humano e sinalização no mercado de trabalho" de Michael Spence, se as sinalizações de um grupo são menos eficazes, acredita-se que haverá menos incentivos a estes trabalhadores de pagarem os custos para emití-las, neste caso, um desincentivo a investir em capital humano.

O efeito negativo da discriminação sobre a economia também vem de encontro com a teoria de esforço-salário justo de [Akerlof e Yellen \(1990\)](#), que afirma que os trabalhadores, baseados em um conceito geral de um valor justo de salário, determina o nível dispersado em uma tarefa de acordo com a relação entre este e o que realmente lhe é pago. A insatisfação devido a injustiça salarial, dentre outros, pelo comportamento discriminatório, tende a reduzir os esforços do trabalhador, gerando desigualdades perniciosas para a economia. Por outro lado, em economias onde o esforço e a produtividade são recompensados e encorajados, tem-se reflexos sobre o crescimento econômico. Isto é, considerando o efeito de um encorajamento sobre os esforços dos trabalhadores, a desigualdade de renda possui uma relação positiva sobre o crescimento ([VOITCHOVSKY, 2005](#)).

A busca por igualdade possui dois princípios: Princípio da Compensação, que defende que diferenças nas circunstâncias devem ser eliminadas e o Princípio da Recompensa, que defende que diferenças no esforço devem ser aceitas. Apesar dos princípios não passarem de dois lados da mesma moeda, a condução de políticas ou estudos que se apoiarem em um desses princípios podem obter resultados conflitantes entre eles ([BRUNORI; FERREIRA; PERAGINE, 2013](#)). Na prática, a distinção destas duas visões surge quando medidas na direção de igualar características de circunstâncias afetam negativamente os resultados para indivíduos com o mesmo esforço ([FLEURBAEY; PERAGINE, 2013](#)). Este resultado pode ser exemplificado através da preocupação presente na literatura da discriminação com relação à políticas afirmativas para grupos, como cotas para negros. Se um lado tendem a corrigir diferenças das características de circunstâncias através do maior acesso à educação (política desejável segundo o princípio da compensação), por outro, pode reduzir o retorno do esforço entre os brancos que despendiam esforços semelhantes ao negro ([COATE; LOURY, 1993](#); [SCHWAB, 1986](#)), que fere o princípio da recompensa liberal, que repudia políticas de redistribuição entre agentes com mesma características.

No modelo de Becker, ([BECKER, 1957](#)) de discriminação por preferências, mesmo sem política ativas de combate ao esse comportamento, o que se observa é que em mercados competitivos com firmas com a mesma função de produção, a discriminação tenderia a ser reduzida por significar custos adicionais. Trabalhadores de grupos minoritários, com salários mais baixos, resultantes da própria discriminação, não seriam contratados pelas firmas que discriminam, o que implicaria em ganhos para empresas que não discriminam devido a maior mão de obra disponível. Contudo, em [Lundberg e Startz \(1998\)](#), utiliza-se do conceito de capital social, observado nos modelos de George J. Borjas, para justificar os efeitos alocativos danosos ao longo do tempo gerados pela discriminação, que origina uma espécie de armadilha da pobreza para grupos de trabalhadores gerando uma persistência intergeracional da discriminação.

<sup>4</sup> Além da própria preferência por discriminar introduzida pelos trabalhos de Becker ([BECKER, 1957](#))

Em [Lundberg e Startz \(1998\)](#) é apresentado um modelo de gerações sobrepostas, de forma a incluir a externalidade do capital humano social de uma geração sobre a outra e a diferenciação deste capital social entre grupos em uma mesma geração. Neste caso, havendo segmentação de grupos na população, o estoque de capital humano influente nas gerações futuras poderá ser relevante apenas dentro destas subpopulações. Este efeito promoveria uma persistência dos níveis de capital humano em cada grupo<sup>5</sup>.

Em muitos trabalhos a evolução e persistência da discriminação é associada a persistência educacional dos pais, apontada como responsável por grande parte da desigualdade de renda ([MALDONADO; MARQUES; FILHO, 2012](#); [SOARES, 2000](#); [CAMPANTE; CRESPO; LEITE, 2004](#); [REIS; RAMOS, 2011](#)). Segundo [BENOIT LOREL \(2001\)](#), a transmissão de escolaridade entre gerações é a chave da desigualdade de escolaridade afetando a desigualdade de renda. A persistência educacional dos pais é uma variável interessante, não apenas pelo efeito direto sobre os descendentes, mas por englobar diversas características de circunstâncias a que estes serão expostos de forma semelhante.

Em [Black \(1995\)](#), em um Modelo de *Search*, a sobrevivência do comportamento discriminatório é contraposto ao tamanho do poder de mercado das firmas, de forma a convergir com o modelo de Becker de discriminação por preferência ([BECKER, 1957](#)). Havendo poucas empresas atuantes, aquelas que discriminam podem usufruir do poder de mercado para pagar seu custo por discriminar. Quando empresas entram, ocorre um aumento na concorrência por trabalhadores e o próprio mercado elimina o excesso de discriminação por tornar custoso a discriminação no momento de contratação no mercado de trabalho.

Também através de um modelo de *Search*, a sobrevivência do comportamento é analisada em [Lundberg e Startz \(2007\)](#), no qual o foco é dado aos retornos dos empregadores ao utilizarem estratégias diferentes de contratação, incluindo níveis diferentes de discriminação. O comportamento discriminatório é modelado como um distanciamento informacional dos empregadores quanto a um grupo específico, forma introduzida em [Lundberg e Startz \(1983\)](#). O tratamento diferenciado entre grupos de trabalhadores poderia ocorrer não pela diferenciação média da produtividade, mas simplesmente por falta de conhecimento a respeito de características deste grupo. Como exemplificado no trabalho citado, supondo que os empregadores se originassem de um destes grupos, isto levaria a uma afinidade maior na interpretação de sinalizações deste mesmo grupo, tornando o outro mais distante informacionalmente. O menor nível de informação levaria a uma diferenciação salarial, mesmo com distribuição de produtividade igual entre os grupos de trabalhadores. Dentro desta estrutura observa-se que grupos minoritários possuem maior propensão a serem excluídos: 1) quanto menor for o grupo; 2) maior a distância informacional destes para com os empregadores; e 3) o custo de procurar trabalhadores é relativamente baixo em relação ao retorno potencial.

### 3 Metodologia

#### 3.1 Variáveis e dados

A base de dados utilizada neste trabalho advém do Instituto Brasileiro de Geografia e Estatística - IBGE, com dados referentes à Pesquisa Nacional por Amostra de Domicílio - PNAD. A princípio os anos selecionados para a análise foram os anos de 1996 e 2014, que possuem um questionário suplementar sobre mobilidade educacional. São informações sobre o nível educacional e ocupação dos pais dos entrevistados, que é utilizado como *proxy* para a origem socioeconômica do agente, permitindo maior controle das características de

<sup>5</sup> Esta persistência tende a se agravar dada possibilidade de mobilidade dos indivíduos. Se dentro o grupo de menor capital humano os indivíduos que obtêm maior acumulação deixam para traz os que possuem menos, resulta em manutenção do baixo capital social deste primeiro grupo.

circunstâncias influentes sobre o salário. As PNADs entre 1996 e 2014 são utilizadas como forma de preencher a análise dinâmica proposta aqui.

A população a que se refere este trabalho foi restrita a homens, moradores de área urbana, com idade entre 20 e 70 anos, com renda na semana da pesquisa maior que zero e mais de 20 horas semanais trabalhadas. Excluiu-se ainda militares e funcionários públicos. São separados dois grupos denominados "não brancos", composto de indivíduos auto-declarados de cor preta ou parda; e os "brancos", sendo indivíduos auto-declarados de cor branca.

A variável de interesse é a diferença na remuneração do trabalho entre os dois grupos. Utilizou-se então (o logaritmo neperiano) a razão da hora-trabalhada no trabalho principal entre "não-brancos" e "brancos". De forma a tornar os valores mais intuitivos, a remuneração foi padronizada em termos mensais, considerando um mês com quatro semanas e 40 horas semanais. Enquanto o vetor de características observáveis é apresentado segundo sua influência na literatura. Sendo:

- Esforço: Nível de escolaridade (anos de estudo ou o grau de instrução), experiência (idade em que começou a trabalhar menos idade atual);
- Temporal: Idade do trabalhador dado ciclo de vida (foram divididas em 4 categorias, sendo: 1) até 21 anos, 2) 21 a 30 anos, 3) 30 a 45 anos, e 4) mais de 45 anos;
- *Background* familiar: Educação do pai, ocupação do pai (em falta, utilizou-se os dados da mãe);
- Posto de Trabalho: Ocupação, se trabalhador formal,
- Espacial: Região do país em que reside, se morador de região metropolitana.

O uso das classificações de ocupações do indivíduo e do pai até 1999 seguem o código do Censo de 1991 e foram aqui classificadas de acordo com [Figueiredo, Silva e Rego \(2012\)](#)<sup>6</sup>, a partir de uma adaptação de Pastore (1979, apud [Figueiredo, Silva e Rego, 2012](#)) que aloca as ocupações declaradas na PNAD de 1996 em seis estratos socioeconômicos. Estes estratos foram definidos segundo um ranqueamento de um índice de status socioeconômico que utiliza informações educacionais e econômicas ([PASTORE; SILVA, 2000](#)).

As classificações e algumas das ocupações englobadas por estes estratos são: Categorias: 1) Alto: grandes proprietários, magistrados, nível superior, etc.; 2) Médio Superior: assessores, gerentes de alto nível, diretores, etc.; 3) Médio Médio: praticistas, viajantes, chefes de nível baixo, etc.; 4) Médio Inferior: pedreiro, eletricitista, marceneiro, etc.; 5) Baixo Superior: braçal, servente, vigias, etc. e; 6) Baixo Inferior: Lavrador, pescador, lenhador, etc.

Após o Censo populacional brasileiro de 2000, houve uma atualização dos códigos das ocupações das PNADs que passaram a seguir a CBO-Domiciliar<sup>7</sup>. De forma a manter a consistência da análise ao longo dos anos, uma nova classificação, seguindo aquela elaborada por [Figueiredo, Silva e Rego \(2012\)](#), foi realizada fazendo uso dos dados para o ano de 2014. O processo de compatibilização baseou-se primeiramente em identificar as ocupações que permaneceram com um mesmo código, de forma a serem alocadas nos mesmos estratos<sup>8</sup>. Estas representaram 42% de toda a pesquisa (74.224 observações). A partir desta alocação inicial utilizou-se o método de análise discriminante para definição das observações restantes.

<sup>6</sup> Agradecimento aos autores pela disponibilidade dos códigos das ocupações estratificadas.

<sup>7</sup> Classificação Domiciliar Brasileira

<sup>8</sup> As compatibilizações dos códigos entre as duas classificações usadas na PNAD foi feita segundo a tabela de correspondência do IBGE disponível em <https://concla.ibge.gov.br/classificacoes/correspondencias/ocupacao-e-posicao-na-ocupacao.html>



### 3.2 Propensity Score Matching

Para a análise das diferenças salariais entre “não brancos” e “brancos” é necessário considerar os diferenciais de características de esforço e circunstâncias existentes entre os grupos. Somente desta forma, é possível destacar a discriminação *pos-market* relacionada à cor da pele. O uso do *Propensity Score Matching* se justifica pela possibilidade de criação de “blocos” de indivíduos semelhantes quanto a estas variáveis, de forma a diferenciá-los apenas pela cor da pele.

De maneira formal, considera-se para um indivíduo  $i \in \{1, \dots, N\}$ , as correspondentes variáveis de interesse  $Y_i(1)$ ,  $Y_i(0)$ , referentes ao resultado obtido ao se receber um determinado tratamento ou não, respectivamente. Para adaptação ao tema em questão, por tratamento, considera-se a cor da pele não ser branca e a variável de interesse os rendimentos do trabalho, ou seja,  $Y_i(1)$  é a remuneração salarial de um trabalhador de cor preta ou parda, “não branco”, e  $Y_i(0)$  é a remuneração deste mesmo trabalhador com a cor da pele branca. O objetivo é mensurar o Impacto Médio do Tratamento sobre os Tratados - ATT<sup>9</sup> em relação a variável de interesse. A discriminação é então observada caso  $Y_i(1) \neq Y_i(0)$ .

Frente a impossibilidade de usar um mesmo indivíduo como contrafactual, faz-se uso de indivíduos diferentes, divididos em grupo de tratamento e de controle, contudo esta pode apresentar um viés de seleção. Formalmente, seja  $T = \{0, 1\}$  uma variável *dummy* que indica se o indivíduo pertence ao grupo de tratamento, então, se

$$E[Y_i(0)|T = 1] - E[Y_j(0)|T = 0] \neq 0, \quad \text{para } i \neq j \quad (1)$$

diferenças na distribuição de características entre os grupos de tratamento e controle já justificariam diferença das médias, independente do tratamento ter ocorrido ou não<sup>10</sup>.

O método de pareamento busca definir um contrafactual mais adequado de forma a eliminar este viés existente. O Método de Pareamento das Características Observáveis encontra e compara indivíduos que sejam semelhantes em termos de um vetor de variáveis observáveis,  $X$ , nos grupos distintos. Formalmente, procura-se um  $X$  que garanta a Condição de Seleção Sobre as Observáveis, *Conditional Independence Assumption - CIA*, dada por:

$$Y_i(0) \perp T|X \Rightarrow E[Y_i(0)|X = x, T = 1] = E[Y_j(0)|X = x, T = 0] \quad (2)$$

Em Rosenbaum e Rubin (1983) foi desenvolvido o método de *Propensity Score Matching* que ao invés de considerar diversas características de comparação, considera apenas uma, um escore de propensão a pertencer ao grupo de tratamento. Este escore de propensão por sua vez é definido pelas variáveis observáveis.

Faz-se  $P(X = x) = Pr(T = 1|X = x)$ , o que implica em uma modificação na hipótese de independência apresentada na Equação 2:

$$Y_i(0) \perp T|P(X) \Rightarrow E[Y_i(0)|P(X = x), T = 1] = E[Y_j(0)|P(X = x), T = 0] \quad (3)$$

A estimativa do *Propensity Score*,  $P(X)$ , é elaborada através de uma Função *Logit* na qual define a probabilidade de um indivíduo participar do grupo de tratamento segundo o vetor de variáveis  $X$ , isto é:

$$P(X = x) = Pr(T_i = 1|X = x) = \frac{\exp(x\beta)}{1 + \exp(x\beta)} \quad (4)$$

<sup>9</sup> Average Treatment Effect on Tread

<sup>10</sup> A variável  $T_i(0)|T = 1$  tem uma implicação teórica. Este refere-se à remuneração de um indivíduo que pertence aos “não brancos” contudo fosse visto como “branco”.



Os indivíduos são então comparados em blocos com aqueles mais próximos no *rank*. Para tanto, ainda deve permanecer a preocupação com as variáveis que serão usadas na estimativa e deve-se atentar para a definição desta proximidade. O método mais comumente utilizado é o de pareamento com o vizinho mais próximo (*Nearest Neighbor Matching*), no qual o indivíduo do grupo de controle mais próximo no ranqueamento através de  $P(x)$  é utilizado como contrafactual para o indivíduo do grupo de tratamento.

De forma a tornar a comparação mais consistente com melhor aproveitamento da amostra, ao invés de utilizar apenas um contrafactual na comparação, utiliza-se do Método de *Kernel Matching* que utiliza todas as unidades de controle do bloco ponderadas inversamente pela distância. Contudo, este método é extremamente custoso em termos de processamento computacional e seu uso nesse trabalho é apenas de indicador de consistência para os resultados obtidos pelo método *Nearest Neighbor Matching*<sup>11</sup>.

Em suma, a proposta de uso deste método para a análise da discriminação, considera o grupo de tratamento como indivíduos que se auto declararam de cor preta ou parda, denominados aqui como grupo dos "não branco" e o grupo de controle, indivíduos auto declarados de cor branca, denominados grupo dos "brancos". O objetivo é determinar a razão entre os rendimento do trabalho de um indivíduo "não branco" em relação ao seu contrafactual, os indivíduos do grupo "branco". Para estabelecer o *score* que determinará os semelhantes, utiliza-se de variáveis de esforço observado, educação e anos de experiência, e controles sobre o *background* familiar, a distribuição espacial e temporal e a segmentação ocupacional através de grupos socioeconômico, baseado na suposição de que tais características sejam distribuídas de forma não-aleatória e possuam efeito sobre a variável de interesse, a remuneração salarial.

### 3.2.1 Análise de sensibilidade: Teste de Robustez

A garantia da CIA<sup>12</sup> para aplicação do pareamento se torna extremamente teórica e, portanto, difícil de ser testada, pois se relaciona com dados não observáveis. Contudo, em Ichino, Mealli e Nannicini (2008) é apresentado uma análise de sensibilidade do resultado ao inserir possíveis variáveis omitidas criadas a partir de parâmetros que são utilizados de forma e verificar a robustez dos resultados.

O teste é baseado na suposição de existência de uma variável binária  $U$ , não observada ou apenas omitida da análise, tal que, sua presença no vetor  $X$  é que garantiria a CIA, isto é:

$$Y_i(0) \perp T/X, U \Rightarrow E[Y_i(0)/X, U, T = 1] = E[Y_j(0)/X, U, T = 0] \quad (5)$$

e, conseqüentemente,

$$E[Y_i(0)/X, T = 1] \neq E[Y_j(0)/X, T = 0]$$

Esta variável é baseada em uma distribuição com parâmetros  $p_{ij}$  que relacionam a probabilidade da sua presença no grupo  $i$  com um resultado  $j$ . Considerando estas variáveis como variáveis binárias, têm-se quatro parâmetros definidores do comportamento da distribuição de  $U$ , para  $i, j \in \{0, 1\}$ :

$$p_{ij} \equiv Pr(U = 1|T = i, Y = j) = Pr(U = 1|T = i, Y = j, X) \quad (6)$$

Na qual,  $i = 1$  indica o grupo de tratamento, "não branco" e  $j = 1$  um resultado sobre a variável de interesse acima de 50% do rendimento da amostra. Neste caso, uma variável com

<sup>11</sup> Segundo Nannicini et al. (2007) todos os métodos possuem estimativas consistentes sob a CIA e a condição de sobreposição.

<sup>12</sup> Conditional Independence Assumption

parâmetro  $p_{10} > p_{00}$ , por exemplo, refere-se a uma possível variável omitida caracterizada pela maior prevalência no grupo de tratamento,  $i = 1$ , que sobre o grupo de controle; dentre aqueles que possuem um rendimento menor que a média,  $j = 0$ .

Apesar destas variáveis apenas apresentarem o efeito positivo e negativo sobre resultado e tratamento, pode-se ainda auferir a magnitude destas influências. Seja as razões das chances, *odds ratio*, de se obter um resultado ou de ser tratado dado  $U = 1$  ou  $U = 0$ , o efeito sobre o resultado é dada por:

$$\Gamma = \frac{\frac{Pr(Y=1|T=0,U=1,W)}{Pr(Y=0|T=0,U=1,W)}}{\frac{Pr(Y=1|T=0,U=0,W)}{Pr(Y=0|T=0,U=0,W)}} \quad (7)$$

Assim como a magnitude do Efeito Seleção da variável em relação ao grupo:

$$\Lambda = \frac{\frac{Pr(T=1|U=1,W)}{Pr(T=0|U=1,W)}}{\frac{Pr(T=1|U=0,W)}{Pr(T=0|U=0,W)}} \quad (8)$$

Cria-se então uma variável controlada, baseada na relação com a variável de resultado e de tratamento. Através de repetidas simulações, a variável criada é utilizada como uma variável observada na configuração dos *matchings* obtendo por fim um ATT médio. Por fim, é comparado o resultado obtido anteriormente à inclusão da variável simulada, verificando a robustez e manutenção da hipótese de Seleção das Observáveis para variáveis com comportamento semelhantes. A aplicação deste método, como sugerida em [Ichino, Mealli e Nannicini \(2008\)](#), é feita a partir de variáveis que “imitem” o comportamento de variáveis observadas já utilizadas no pareamento original. Este processo permite obter uma intuição por traz das variáveis artificiais criadas.

A utilização da análise de *Propensity Score* é feita através do Programa Estatístico Stata 12.0, seguindo as rotinas desenvolvidas em [Abadie et al. \(2004\)](#) e [Becker, Ichino et al. \(2002\)](#). As análises de sensibilidades são apresentadas em [Ichino, Mealli e Nannicini \(2008\)](#) e [Nannicini et al. \(2007\)](#) para o Stata através da função SENSATT.

## 4 Resultados e discussões

### 4.1 Evolução das características entre "brancos" e "não brancos" no mercado de trabalho brasileiro entre os anos de 1996 e 2014

A Tabela 1 apresenta uma síntese de informações referentes à educação e salário para os dois grupos analisados, de forma a destacar as diferenças existentes e como estas evoluíram ao longo do período. Vale lembrar que a amostra foi restrita a homens, moradores de área urbana, com idade entre 20 e 70 anos, com renda na semana da pesquisa maior que zero e com mais 20 horas semanais trabalhada. Excluiu-se ainda da amostra militares e funcionários públicos. Já o salário médio, trata-se da remuneração média da hora-trabalho, do trabalho principal, padronizada para 40 horas semanais em um mês com 4 semanas. É apresentado ainda de forma deflacionada através do Índice de Preço ao Consumidor Amplo - IPCA, a preços de Janeiro de 2015.

Observa-se que em 1996, o salário médio dos "não-brancos" era de R\$ 1096,31 (a preços de 2015) o que representava 48,6% do valor médio recebido entre os "brancos". Em 2014, essa razão era de 59,9%. Essa menor diferença de ganhos é resultados de uma elevação de 36% no ganho real salarial dos "não brancos", contra um ganho de pouco mais de 10% para o outro grupo.

Parte da diferença salarial, assim como a evolução desta, pode ser explicada pela configuração educacional dos trabalhadores de cada grupo. Segundo a Tabela 1, em 2014 a

Tabela 1 – Dados médios anuais por grupo entre 1996 e 2014

ano	cor	% pop.	Média de anos de estudo	% com + 15 anos de estudo	Salário médio (padronizado)	Razão (salário)
1996	ñ-brancos	40,2%	5,1	2,1%	R\$ 1.096,31	48,6%
	brancos	59,8%	7,3	9,6%	R\$ 2.257,29	
1997	ñ-brancos	41,4%	5,2	2,3%	R\$ 1.073,52	47,9%
	brancos	58,6%	7,5	9,8%	R\$ 2.240,55	
1998	ñ-brancos	41,3%	5,3	2,1%	R\$ 1.064,12	48,1%
	brancos	58,7%	7,6	10,6%	R\$ 2.211,62	
1999	ñ-brancos	41,7%	5,4	2,2%	R\$ 991,20	49,4%
	brancos	58,3%	7,7	10,5%	R\$ 2.005,69	
2001	ñ-brancos	43,4%	5,7	2,3%	R\$ 952,60	49,2%
	brancos	56,7%	7,9	10,1%	R\$ 1.934,99	
2002	ñ-brancos	43,8%	5,9	2,5%	R\$ 969,09	50,5%
	brancos	56,2%	8,1	10,4%	R\$ 1.918,99	
2003	ñ-brancos	44,7%	6,1	2,5%	R\$ 892,29	49,0%
	brancos	55,3%	8,3	10,9%	R\$ 1.822,72	
2004	ñ-brancos	45,3%	6,4	2,7%	R\$ 930,52	51,9%
	brancos	54,7%	8,4	10,8%	R\$ 1.792,80	
2005	ñ-brancos	46,7%	6,6	2,9%	R\$ 966,22	52,2%
	brancos	53,3%	8,6	11,3%	R\$ 1.851,42	
2006	ñ-brancos	47,2%	6,8	3,1%	R\$ 1.037,61	52,5%
	brancos	52,8%	8,8	11,9%	R\$ 1.975,29	
2007	ñ-brancos	48,0%	7,0	3,3%	R\$ 1.085,57	53,7%
	brancos	52,0%	8,9	12,8%	R\$ 2.020,04	
2008	ñ-brancos	49,2%	7,2	3,9%	R\$ 1.146,76	55,3%
	brancos	50,8%	9,1	13,3%	R\$ 2.074,33	
2009	ñ-brancos	49,4%	7,4	4,2%	R\$ 1.211,52	58,1%
	brancos	50,6%	9,2	13,7%	R\$ 2.085,46	
2011	ñ-brancos	50,7%	7,6	4,8%	R\$ 1.325,19	59,6%
	brancos	49,3%	9,3	14,1%	R\$ 2.225,08	
2012	ñ-brancos	51,9%	7,8	5,1%	R\$ 1.409,53	58,1%
	brancos	48,1%	9,5	15,3%	R\$ 2.427,02	
2013	ñ-brancos	52,2%	7,9	5,5%	R\$ 1.467,22	59,2%
	brancos	47,8%	9,6	16,1%	R\$ 2.480,24	
2014	ñ-brancos	53,4%	8,0	5,9%	R\$ 1.490,71	59,9%
	brancos	46,6%	9,7	16,1%	R\$ 2.490,34	

média de anos de estudo entre os “não-brancos” foi de 8 ano, sendo que 5,9% apresentavam mais de 15 anos de estudo. Para os “brancos”, a média é quase 1,8 anos maior, sendo que 16% destes possuíam mais de 15 anos de estudo.

Na Figura 1 a evolução da diferença na escolaridade média ao longo do período é melhor ilustrada. As curvas ascendentes apresentam a escolaridade média enquanto as curvas descendentes (com representação no eixo vertical a direita) apresentam a queda no diferencial desta escolaridade média. Novamente, baseado no esperado impacto da escolaridade sobre a produtividade, parte da redução no diferencial salarial entre “brancos” e “não brancos” tem explicação na queda do diferencial educacional entre os grupos<sup>13</sup>.

A figura 1 compara ainda a amostra da população denominada “Restrita”, analisada até então, e uma “Geral” composta por homens com idade entre 20 e 70 anos, sem restringir as observações por características relativas à participação no mercado de trabalho. Esta análise permite vislumbrar a relação entre uma possível maior demanda do mercado por trabalhadores “não brancos” qualificados e uma melhora no nível educacional dos “não brancos” na população.

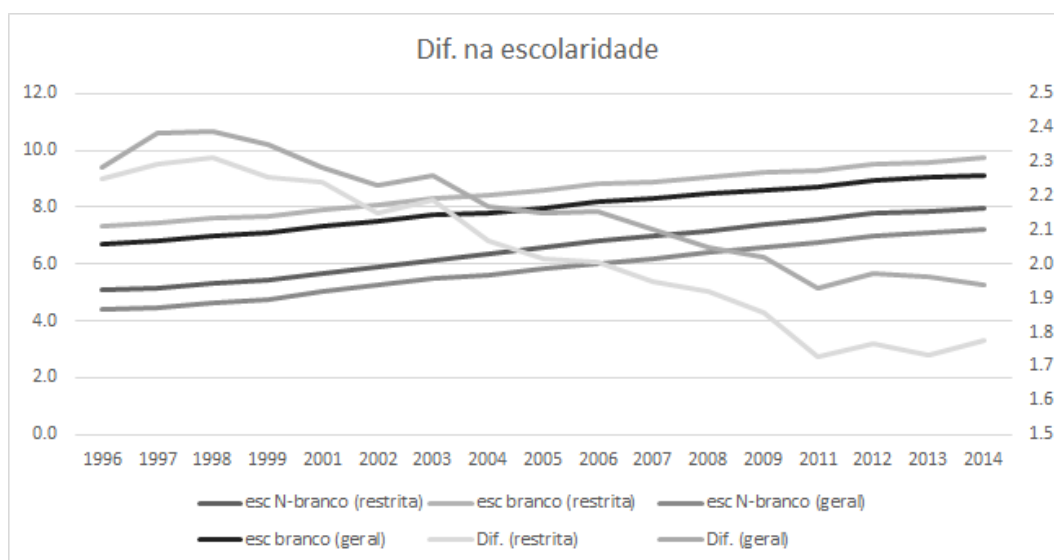


Figura 1 – Dados da diferença da média da escolaridade dos grupos entre 1996 a 2014

Além do maior nível educacional na amostra restrita, a diferença no nível educacional entre os grupos, representada pelas duas linhas declinantes, indica uma diferença menor no nível educacional entre os grupos na amostra “restrita”, que na população da amostra “geral”. Este comportamento estaria de acordo com uma lógica de seleção do mercado pelos mais qualificados. Contudo, também se observa na amostra “restrita” uma queda mais intensa deste diferencial educacional, indicando uma transformação maior no mercado que na população. Tal fato sugere uma melhoria, em termos de igualdade de seleção no *pos-market* comparado a uma melhoria no acesso à qualificação que ocorre fora do mercado de trabalho. Contudo, faz-se necessário investigação mais aprofundada sobre este tipo de discriminação, que não é o cerne deste trabalho.

<sup>13</sup> Diferentemente dos resultados de Silva, França e Neto (2016), em que a elevação na escolaridade poderia ter efeitos de elevação na desigualdade, ao se analisar a desigualdade existente pela cor da pele, este “paradoxo do progresso” não faz sentido. Maior mão de obra qualificada no grupo minoritário tende a contribuir para a redução das diferenças de renda. Em Silva, França e Neto (2016) é constatado ainda uma queda no preço da educação. É possível que em parte, esse “efeito preço”, seja explicado pela maior participação de mão de obra qualificada advinda de um grupo que recebe salários mais baixos. Resultado esperado sob a hipótese de discriminação de mão de obra.

Assim como há diferença na distribuição da escolaridade entre os grupos, outros fatores correlacionados aos ganhos salariais também justificam diferenças dos ganhos médios e na evolução destes entre os grupos. Na Tabela 2 são apresentadas informações quanto a distribuição espacial, etária, educacional e ocupacional dentro dos dois grupos de análise.

Tabela 2 – Dispersão das características dentro de cada grupo

		1996		2014	
		Brancos	ñ-brancos	Brancos	ñ-brancos
		100%	100%	100%	100%
<b>Região</b>	Norte	2,5%	8,8%	3,3%	10,3%
	Nordeste	10,7%	35,1%	12,5%	29,8%
	Sudeste	57,4%	41,2%	52,3%	42,7%
	Sul	24,0%	5,9%	24,9%	7,4%
	Centro	5,5%	9,1%	7,0%	9,8%
<b>Região Metropolitana</b>	sim	38,0%	35,6%	36,2%	35,1%
	não	62,0%	64,4%	63,8%	64,9%
<b>Escolaridade</b>	sem instrução	7,2%	19,4%	3,1%	7,3%
	Fundamental (incomp.)	54,2%	61,7%	21,4%	31,7%
	Fundamental	9,2%	6,6%	16,1%	19,6%
	Médio	19,1%	10,3%	41,8%	35,5%
	Superior	10,3%	2,0%	17,6%	6,0%
<b>Idade</b>	21 a 29 anos	16,8%	20,3%	24,5%	27,4%
	30 a 45 anos	52,5%	51,0%	41,6%	43,4%
	46 a 70 anos	30,7%	28,7%	33,9%	29,2%
<b>Grupo ocupacional socioeconômico</b>	Alto	7,1%	1,9%	5,5%	1,5%
	Médio superior	12,1%	5,1%	12,6%	5,6%
	Médio	12,3%	7,5%	16,3%	10,1%
	Médio inferior	48,1%	50,1%	46,1%	52,1%
	Baixo superior	15,9%	25,2%	15,9%	25,0%
	Baixo inferior	4,6%	10,7%	3,6%	5,7%

Destaca-se da Tabela 2, a maior concentração entre os "não-brancos" na região Nordeste, com redução desta participação enquanto, entre os "brancos", observa-se movimento inverso. Apenas 10,7% dos "brancos" se encontravam na Região Nordeste do país em 1996, passando a 12,5% em 2014; enquanto para o grupo dos "não-brancos" o percentual era de 35,1% em 1996, se reduzindo para 29,8% em 2014. Segundo dados da própria PNAD, a região nordeste é a região brasileira com menores valores de renda média do trabalho (padronizada). Portanto, esta informação reforça o resultado esperado de desigualdade de renda entre os dois grupos assim como a redução no diferencial ao longo do tempo.

Análise semelhante também ocorre quanto aos grupos ocupacionais. Assim como nos níveis educacionais, observa-se a prevalência do grupo dos "não brancos" nos grupos ocupacionais mais baixos enquanto o grupo dos "brancos" possuem maior prevalência nos grupos socioeconômicos superiores.

## 4.2 Análise da diferença média dos ganhos utilizando o PSM

Na tabela 3, novamente é apresentado as razões dos ganhos entre os "não brancos" e "brancos", contudo estes são analisados agora segundo os mesmos níveis de escolaridade, de ocupação ou região.

Como na Tabela 1 a primeira linha apresenta o valor médio pago pelo trabalho, que para 2014 era cerca de R\$1490,00, o que representa 59,9% da renda média do trabalhador "branco" que recebia cerca de R\$2500,00<sup>14</sup>. Ao se restringir a comparação a trabalhadores com os mesmos níveis educacionais, por exemplo, o que se observa é que a diferença de ganhos persiste porém em menor patamar. Novamente, este se justifica pela desigualdade na distribuição dessas características entre os grupos, tal como evidenciada na Tabela 2, sendo elas fatores relevantes na determinação salarial.

Tabela 3 – Renda média entre os grupos segundo características espaciais, educacionais e ocupacionais

	1996			2014		
	não-branco	branco	razão	não-branco	branco	razão
<b>Geral</b>	R\$ 1,096.81	R\$ 2,252.10	48.7%	R\$ 1,490.17	R\$ 2,489.41	59.9%
<b>Por região</b>						
Norte	R\$ 1,111.63	R\$ 1,851.14	60.1%	R\$ 1,385.09	R\$ 1,955.14	70.8%
Nordeste	R\$ 894.54	R\$ 1,730.38	51.7%	R\$ 1,212.91	R\$ 1,662.99	72.9%
Sudeste	R\$ 1,248.45	R\$ 2,466.63	50.6%	R\$ 1,624.45	R\$ 2,756.99	58.9%
Sul	R\$ 1,009.90	R\$ 2,008.52	50.3%	R\$ 1,584.17	R\$ 2,364.57	67.0%
Centro	R\$ 1,202.40	R\$ 2,174.86	55.3%	R\$ 1,791.46	R\$ 2,654.24	67.5%
<b>Por escolaridade</b>						
Sem instrução	R\$ 690.94	R\$ 978.19	70.6%	R\$ 980.75	R\$ 1,213.65	80.8%
Fund. (incomp.)	R\$ 975.37	R\$ 1,409.51	69.2%	R\$ 1,192.73	R\$ 1,436.31	83.0%
Fundamental	R\$ 1,268.34	R\$ 1,849.06	68.6%	R\$ 1,313.81	R\$ 1,629.46	80.6%
Médio	R\$ 1,855.65	R\$ 2,821.76	65.8%	R\$ 1,628.56	R\$ 2,188.86	74.4%
Superior	R\$ 4,996.01	R\$ 6,998.42	71.4%	R\$ 3,825.87	R\$ 5,664.85	67.5%
<b>Por grupo socio-econ.</b>						
Alto	R\$ 5,039.05	R\$ 7,059.24	71.4%	R\$ 5,181.49	R\$ 6,690.23	77.4%
Médio superior	R\$ 2,201.77	R\$ 4,246.69	51.8%	R\$ 3,077.14	R\$ 4,770.46	64.5%
Médio	R\$ 2,072.36	R\$ 3,023.92	68.5%	R\$ 2,184.72	R\$ 3,014.64	72.5%
Médio inferior	R\$ 1,105.17	R\$ 1,580.88	69.9%	R\$ 1,408.87	R\$ 1,701.89	82.8%
Baixo superior	R\$ 677.02	R\$ 966.71	70.0%	R\$ 1,002.90	R\$ 1,346.37	74.5%
Baixo inferior	R\$ 450.38	R\$ 645.93	69.7%	R\$ 833.07	R\$ 1,390.29	59.9%

Como parte do método de *Propensity Score Matching* - PSM proposto para análise da razão dos ganhos entre os grupos, estimou-se a regressão logística, que indicou a probabilidade de um indivíduo com determinada característica pertencer ao grupo "não branco". Destes, gerou-se os *scores* utilizados na criação dos "blocos" de indivíduos com distribuição "balanceada" das características de controle entre os dois grupos. Para o melhor balanceamento a estimação usou das variáveis ligadas ao local de residências (região, região metro-

<sup>14</sup> Valores corrigidos pelo IPCA para janeiro de 2015.



politana por estado), níveis de escolaridade, experiência, grupo ocupacional e se trabalhador formal.

Os resultados seguem a lógica apresentada na Tabela 3, contudo, ao invés de igualar apenas uma característica, como o nível de escolaridade, utilizou-se dos *scores* estimados pelo modelo Logit. O resultado da estimação é apresentado na Tabela 4, na qual constam os dois métodos de definição dos contrafactuais, *Nearest Neighbor Matching* e *Kernel Matching*, citados na metodologia.

Tabela 4 – Propensity Score Matching - 1996 e 2014

Nearest Neighbor Matching								Kernel Matching	
	Att		Razão	Int. conf. 95%	tratamento	controle		Att	Razão
<b>2014</b>	-0.059	0.009	<b>94.3%</b>	0.930%	0.955%	25637	14553	-0.075	92.8%
<b>1996</b>	-0.105	0.017	<b>90.0%</b>	0.876%	0.925%	10001	7466	-0.146	86.4%

Observa-se que a razão dos ganhos entre "não brancos" e "brancos" em 1996 era de 90% (86,4% para o método de Kernel) e para 2014 esse valor se situava em torno de 94% (92,8%), valores superiores àqueles apresentados na Tabela 3, em que essas razões eram de 48,7% em 1996 e 59,9% em 2014. Ou seja, indivíduos de diferentes grupos ainda apresentam diferenciações nas médias salariais, contudo esta se apresenta em menores patamares.

Esses dados revelam dois fatos sobre a discriminação no mercado de trabalho brasileiro. Primeiramente, ressalta-se a menor diferença de ganhos quando fatores relevantes sobre a remuneração são adicionados na análise. O resultado sugere que grande parte das diferenças das remunerações são resultantes de uma discriminação pré-mercado, que leva à média da produtividade de um grupo ser maior e, portanto, a ter uma renda média mais elevada, fato já consolidado nas pesquisas sobre discriminação.

O outro fator é relacionado a evolução temporal desta discriminação. Assumindo que a razão da renda menor que um, encontrada pelo método PSM, se refira à discriminação do mercado de trabalho ligada à cor da pele, esta se reduziu menos que aquela apontada na Tabela 3. Esta evolução é pontualmente observada na Figura 2, que apresenta as razões médias dos ganhos dos "não brancos" em relação aos "brancos" para os anos de 1996 a 2014<sup>15</sup>.

Apesar da tendência de crescimento das razões, o que implica em menores diferenças salariais entre os grupos, observa-se que de 2002 a 2014, a razão foi de 0,933% para 0,942%, menos de 1 ponto percentual. A razão dos ganhos médios entre os grupos neste mesmo período foi de 50,5% para 59,9% segundo a análise exposta na tabela 1.

#### 4.3 Análise de consistência através de uma análise de sensibilidade frente a possíveis variáveis omitidas

Como forma de analisar inicialmente a consistência dos resultados obtidos, são estimados, na Tabela 5, outros dois modelos de PSM: 1) usando apenas variáveis de esforço; 2) que inclui, além das variáveis do modelo inicial, o *background* familiar.

O primeiro modelo permite observar o efeito da omissão de variáveis espaciais e da característica do trabalho, relevantes na determinação dos ganhos e na seleção dos grupos. A diferença na razão dos ganhos médios entre os dois grupos, em relação ao modelo inicial,

<sup>15</sup> As estimações foram feitas através do método *Nearest Neighbor Matching* com intervalos de confiança a 95% de significância. Utilizou-se 400 replicações para obtenção do intervalo de confiança. Seguiu-se as especificações realizadas e já comentadas anteriormente para os anos de 2014 e de 1996.

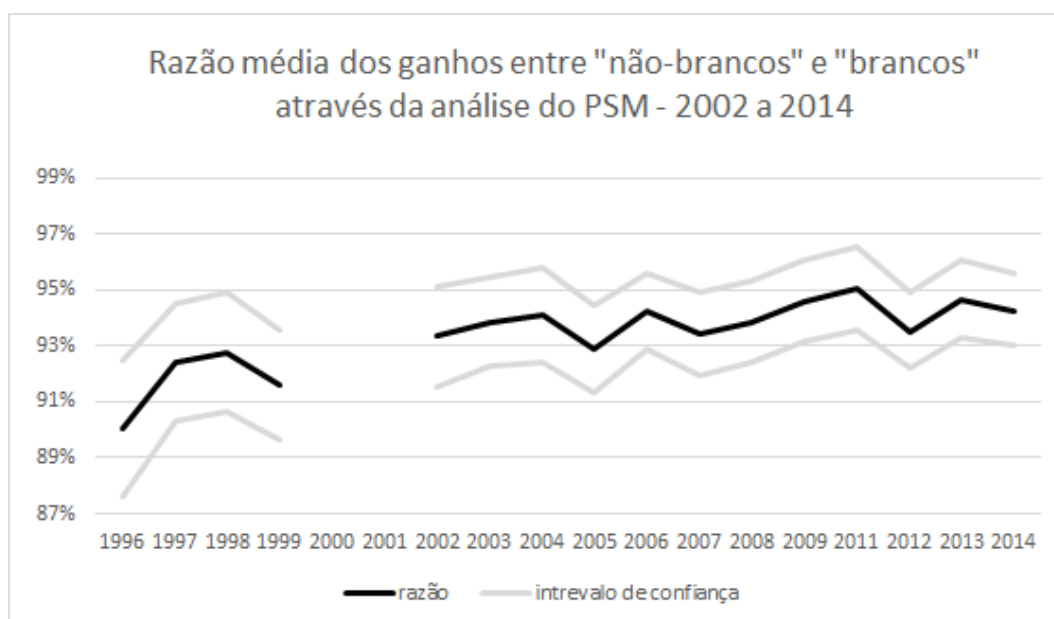


Figura 2 – Razão média dos ganhos entre "não-brancos" e "brancos" através da análise do PSM - 2002 a 2014

Tabela 5 – Propensity Score Matching - 1996 e 2014 - Modelos para testar consistência

		Nearest Neighbor Matching					Kernel Matching		
		Modelo apenas com variáveis de esforço							
	Att	Razão	Int. conf. 95%	tratamento	controle	Att	Razão		
<b>2014</b>	-0.193	0.006	82.4%	36869	28119	-0.212	80.9%		
<b>1996</b>	-0.317	0.011	72.8%	15784	20783	-0.339	71.2%		
		Modelo com <i>background</i> familiar							
	Att	Razão	Int. conf. 95%	tratamento	controle	Att	Razão		
<b>2014</b>	-0.009	99.1%		4441	1876	-0.038	96.3%		
<b>1996</b>	-0.094	0.019	91.0%	0.882%	0.938%	8277	5509	-0.142	86.8%

é de cerca de 12 pontos percentuais para o ano de 2014. Para 1996 essa diferença entre modelos é de cerca de 15 pontos. Observa-se ainda a consistência dessa estimaco com os resultados na Tabela 3 quando controlado apenas os nveis de escolaridade. A razo dos ganhos se situava em torno dos 70% para 1996 e 80% em 2014.

Para o "Modelo com *background* familiar" explorou-se no apenas o estrato socioeconmico do qual o indivduo se encontra, mas tambm de qual estrato ele veio. Fez-se uso de interaes entre variveis ocupacionais do indivduo e de seus pais<sup>16</sup>, que propiciaram, alm de grupos de pareamento bastante homogneos, uma intuio analtica interessante. Criou-se *dummies* para cada nvel social de ocupao do pai e filho, por exemplo, a varivel "ocup1\_12" indica um indivduo que pertence ao estrato "Alto" ou "Mdio Superior" e que possui pais no estrato "Alto", como definidos na Tabela 9 no anexo A.

O resultado deste ltimo modelo  que em 1996 as razes de ganho estatisticamente no se diferenciaram, variao na mdia em torno de 1%, indicando consistncia da estima-

<sup>16</sup> Utilizou-se das caractersticas ocupacionais dos pais, advindas das pesquisas suplementares da PNAD de 1996 e 2014.

ção inicial quanto ao viés que poderia surgir quanto ao *background* familiar. Para 2014, há um grande número de observações sem informação quanto a ocupação dos pais. A amostra que era de mais de 40.000 observações cai para pouco mais de 6.000, o que resultou em uma amostra viesada. Ao ser retirada as observações sem informação, a amostra apresenta resultados de renda muito próximos entre os dois grupos, mesmo sem considerar controles do *background* familiar, educação e variáveis espaciais. Portanto, a estratégia do uso desse suplemento falha para 2014 e não traz características dos pais como uma possível variáveis omitidas para 1996.

É válido lembrar que os resultados encontrados baseiam-se na inexistência de variáveis não observadas ou omitidas que influenciem o resultado e seleção do grupo. Fazendo uso do método desenvolvido em Nannicini et al. (2007). Busca-se testar a robustez do resultado para 2014. Na estimação do modelo inicial do PSM, foi inseridas possíveis variáveis omitidas que “imitam”, em termos de distribuição, o comportamento de variáveis relacionadas à condição educacional e ocupacional dos pais em 1996. Ou seja, como não se tem informação confiável quanto a estas características para 2014, utilizou-se uma simulação a partir de características do passado. São explorados dois conjuntos de variáveis: Aquelas caracterizadas pela maior prevalência no grupo de controle, "brancos", e que também estejam correlacionadas com os melhores resultados. Neste caso, esta variável, e não a discriminação da cor em si, poderia ser a responsável pelo melhor desempenho do grupo de controle, por exemplo, educação de nível superior. E o outro conjunto de variáveis, caracterizado pela correlação com piores rendimentos, porém mais concentrada no grupo de tratamento, "não brancos", gerando um efeito que justificaria o pior resultado deste.

Formalmente, estes dois conjuntos de variáveis são respectivamente caracterizados por uma variável com os seguintes parâmetros:

$$p_{01} > p_{00} \text{ e } p_1 < p_0 \text{ (} d > 0 \text{ e } s < 0 \text{) ou}$$

$$p_{01} < p_{00} \text{ e } p_1 > p_0 \text{ (} d < 0 \text{ e } s > 0 \text{)}$$

A tabela 6 apresenta os parâmetros e resultados da análise<sup>17</sup>. A primeira linha apresenta o valor do ATT encontrado inicialmente, seguindo-se das possíveis variáveis omitidas.

Tabela 6 – **Análise de sensibilidade para 2014, com efeito resultado > 1 e um efeito seleção < 1 ( d > 0 e s < 0)**

	p11	p10	p01	p00	Resultado	Seleção	Pr(U)	Razão
PSM (original)	—	—	—	—	—	—	—	<b>94.2%</b>
pais_ocup_baixo (inf)_1996	18%	28%	14%	26%	0.466	1.228	22.3%	<b>94.6%</b>
pais_esc<=fund_inc_1996	76%	82%	73%	84%	0.514	1.044	79.0%	<b>94.6%</b>
esc_fund_incompleto_1996	54%	65%	42%	70%	0.309	1.174	58.2%	<b>95.1%</b>
ocup_baixo_1996	12%	46%	8%	38%	0.142	1.492	28.6%	<b>95.9%</b>
pais_ocup>=medio_1996	7%	3%	13%	5%	2.9	0.5	6.6%	<b>94.8%</b>
pais_esc>=medio_1996	6%	1%	13%	2%	7.4	0.4	5.1%	<b>95.4%</b>
pais_esc_superior_1996	2%	0%	6%	1%	6.4	0.2	2.0%	<b>95.4%</b>
esc_superior_1996	6%	0%	17%	1%	21.0	0.3	5.5%	<b>96.2%</b>
esc>=medio_1996	29%	6%	44%	9%	8.0	0.5	20.3%	<b>97.2%</b>

<sup>17</sup> Selecionou-se dentre as variáveis inicialmente utilizadas aquelas que possuem parâmetros mais influenciáveis sobre o tratamento.

Segundo a Tabela 6 a inserção das possíveis variáveis omitidas não foram capazes de anular o Impacto negativo de pertencer ao grupo "não-branco" sobre os rendimentos, representado na razão dos rendimentos entre os grupos. Em termos de ocupação e escolaridade dos pais, a variável que traz um resultado mais distante do encontrado anteriormente é a variável "pais\_esc\_superior\_1996" que imita a distribuição do ensino superior entre os pais nos dois grupos e sobre os resultados para 1996. Neste caso, a variável possui uma prevalência de apenas 2% na amostra analisada, contudo seu efeito seleção é de 0,2 (isto é, para cada 100 indivíduo dentro o grupo "branco" com ensino superior, observa-se 20 no grupo "não-branco" com essa mesma escolaridade); e um efeito resultado de 6,4 (ou seja, para cada 100 indivíduo com ensino superior dentre os que recebem rendimentos menores que a média, observa-se mais de 640 com iguais características dentre os que recebem mais que a média). Ou seja, é uma variável com significativa presença entre aqueles que recebem mais altos rendimentos e concentrada entre os "brancos".

Dado significativa relevância das variáveis educacionais, apresentou-se ainda algumas variáveis que imitam a distribuição educacional de 1996. A variável relacionada à educação com maior impacto foi a "esc $\geq$ medio\_1996" que indica uma distribuição do nível educacional maior ou igual ao ensino médio dentro da amostra de 1996. Esta variável "criada" reduziu a diferença entre o rendimento médio dos grupos, elevando a razão de ganho dos "não-brancos" para 97,2% do que recebem os "brancos". Essa variável, além de trazer uma forte correlação positiva sobre o rendimento (por imitar o impacto da educação), possui um apelo crítico nos valores de seus parâmetros. Carrega em si uma menor disseminação da educação formal na população em 1996 e, conseqüente maior efeito diploma do passado; e uma maior concentração no grupo dos brancos, considerando as melhoras distributivas da educação ao longo dos anos.

Em suma, as estimações da Tabela 6 sugerem consistência dos resultados iniciais, no sentido de que a maior alteração obtida no resultado original, de quase dois e meio pontos percentuais, exigiria a omissão de uma variável com parâmetros extremos, o que a torna mais difícil de realmente existir, e mesmo assim não sendo capaz de eliminar a diferença que aqui é relacionada com a discriminação pela cor da pele.

Em anexo encontram-se as Tabelas 7 e 8 que apresentam a mesma análise para outras possíveis variáveis omitidas que imitam variáveis de 2014. A maior variação do impacto ocorre com a inclusão da variável que apresenta parâmetros semelhantes à variável "esc\_superior", referente ao indivíduo possuir educação superior. Novamente uma variável educacional. A razão de ganhos de 94,2% passa para 96,6%, alterando menos que as mesmas variáveis baseadas na distribuição de 1996.

Por fim, vale citar o resultado da análise de sensibilidade para o ATT, estimados para o ano de 1996. O impacto das variáveis imitadas seguem o mesmo padrão observado para a análise de 2014, onde a maior diferenciação ocorre novamente com uma variável omitida ligada ao nível educacional, que proporcionou a variação de 91% para 94,2% na razão média dos ganhos, tanto para as variáveis educacionais mais elevadas ( $d > 0$  e  $s < 0$ ) quanto para as mais baixas ( $d < 0$  e  $s > 0$ ), contudo, nos demais resultados, inclusive ligados às características dos pais, observa-se grande consistência em relação ao resultado original.

## 5 Considerações finais

Este trabalho buscou caracterizar a discriminação salarial no Brasil motivada pela cor da pele do trabalhador. As estimativas ao longo do trabalho foram testadas pelo uso de diferentes métodos de pareamento, pela análise de sensibilidade dos resultados frente a possível vies oculto, além da própria estimativa ao longo dos anos que não mostrou trajetória incoerente, sem grandes saltos, trazendo credibilidade aos resultados encontrados no que tange

sua consistência. Da discussão frente aos resultados, três pontos principais se destacam. Primeiramente, torna-se evidente a existência de uma discriminação pela cor no mercado de trabalho brasileiro ao longo das duas últimas décadas. Esta se evidencia amparada na literatura consultada e na impossibilidade de obter ganhos idênticos entre "brancos" e "não-brancos" a partir da abordagem do método de *Propensity Score Matching*, mesmo considerando variáveis sintéticas extremas, não conhecidas, com potenciais geradoras de desigualdade.

Um segundo ponto se relaciona ao tamanho dessa discriminação. A razão salarial existente entre "brancos" e "não brancos" está em torno de 59%, segundo os dados da PNAD. Contudo, este valor deve ser desvinculada da ideia de mensuração do tamanho da discriminação. Quando consideradas as diferenças de distribuição de características produtivas e espaciais entre estes dois grupos para o ano de 2014, a razão dos ganhos ficou em torno de 94%. Portanto se acredita que haja um perda salarial de 6% para os "não brancos" devido a sua cor, e não os 41% sugeridos inicialmente.

Por fim, quanto à persistência do comportamento discriminatório, os resultados na Figura 2 indicam, principalmente entre 2002 e 2014, uma estabilidade na presença deste comportamento no mercado de trabalho. Na média, houve uma variação de 1% (de 0.933% em 2002, para 0.942% em 2014). O que leva a uma mudança de percepção no combate à discriminação se comparado com a melhoria de quase dez pontos percentuais na média de ganhos entre os dois grupos (de 50,5% em 2002, para 59,9% em 2014) que em grande parte advém da melhora na distribuição de características entre os grupos ao longo do período.

## Referências

- ABADIE, A. et al. Implementing matching estimators for average treatment effects in stata. *Stata journal*, STATA PRESS, v. 4, p. 290–311, 2004. 8
- AKERLOF, G. A.; YELLEN, J. L. The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics*, JSTOR, p. 255–283, 1990. 3
- ARROW, K. The theory of discrimination. *Discrimination in labor markets*, Princeton, v. 3, n. 10, p. 3–33, 1973. 1, 2, 3
- BECKER, G. S. The economics of discrimination. *University of Chicago Press*, 1957. 1, 3, 4
- BECKER, S. O.; ICHINO, A. et al. Estimation of average treatment effects based on propensity scores. *The stata journal*, v. 2, n. 4, p. 358–377, 2002. 8
- BLACK, D. A. Discrimination in an equilibrium search model. *Journal of labor Economics*, JSTOR, p. 309–334, 1995. 4
- BRUNORI, P.; FERREIRA, F. H.; PERAGINE, V. Inequality of opportunity, income inequality and economic mobility: some international comparisons. *World Bank Policy Research Working Paper*, n. 6304, 2013. 3
- CAMPANTE, F. R.; CRESPO, A. R.; LEITE, P. G. Desigualdade salarial entre raças no mercado de trabalho urbano brasileiro: aspectos regionais. *Revista Brasileira de Economia*, SciELO Brasil, v. 58, n. 2, p. 185–210, 2004. 4

- COATE, S.; LOURY, G. Antidiscrimination enforcement and the problem of patronization. *The American Economic Review*, JSTOR, p. 92–98, 1993. 1, 3
- FIGUEIREDO, E. A. d.; SILVA, C. R. d. F.; REGO, H. d. O. Desigualdade de oportunidades no brasil: efeitos diretos e indiretos. *Economia Aplicada*, SciELO Brasil, v. 16, n. 2, p. 237–254, 2012. 5
- FLEURBAEY, M.; PERAGINE, V. Ex ante versus ex post equality of opportunity. *Economica*, Wiley Online Library, v. 80, n. 317, p. 118–130, 2013. 3
- ICHINO, A.; MEALLI, F.; NANNICINI, T. From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics*, Wiley Online Library, v. 23, n. 3, p. 305–327, 2008. 7, 8
- LUNDBERG, S.; STARTZ, R. On the persistence of racial inequality. *Journal of Labor Economics*, The University of Chicago Press, v. 16, n. 2, p. 292 – 323, Abril 1998. 2, 3, 4
- LUNDBERG, S.; STARTZ, R. Information and racial exclusion. *Journal of Population Economics*, Springer, v. 20, n. 3, p. 621–642, 2007. 4
- LUNDBERG, S. J.; STARTZ, R. Private discrimination and social intervention in competitive labor market. *The American Economic Review*, JSTOR, p. 340–347, 1983. 3, 4
- MALDONADO, W. L.; MARQUES, I. M.; FILHO, O. C. d. S. A dynamic model of education level choice: application to brazilian states. *Revista Brasileira de Economia*, SciELO Brasil, v. 66, n. 2, p. 225–245, 2012. 4
- NANNICINI, T. et al. Simulation-based sensitivity analysis for matching estimators. *Stata Journal*, Citeseer, v. 7, n. 3, p. 334, 2007. 7, 8, 15
- PASTORE, J.; SILVA, N. do V. *Mobilidade social no Brasil*. [S.l.]: Makron Books, 2000. 5
- PHELPS, E. S. The statistical theory of racism and sexism. *The american economic review*, JSTOR, p. 659–661, 1972. 1, 2
- REIS, M. C.; RAMOS, L. Escolaridade dos pais, desempenho no mercado de trabalho e desigualdade de rendimentos. *Revista Brasileira de Economia*, SciELO Brasil, v. 65, n. 2, p. 177–205, 2011. 4
- ROEMER, J. E. Equality of opportunity: A progress report. *Social Choice and Welfare*, Springer, v. 19, n. 2, p. 455–471, 2002. 2
- 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. 6
- SCHWAB, S. Is statistical discrimination efficient? *The American Economic Review*, JSTOR, p. 228–234, 1986. 3
- SILVA, V. H. M. C.; FRANÇA, J. M. S. d.; NETO, V. R. d. P. Capital humano e desigualdade salarial no brasil: uma análise de decomposição para o período 1995-2014. *Estudos Econômicos (São Paulo)*, SciELO Brasil, v. 46, n. 3, p. 579–608, 2016. 10
- SOARES, S. S. D. O perfil da discriminação no mercado de trabalho. *Homens negros, mulheres brancas e mulheres negras. Estudos Sociais do IPEA, Texto para discussão*, n. 769, 2000. 4
- VOITCHOVSKY, S. Does the profile of income inequality matter for economic growth? *Journal of Economic Growth*, Springer, v. 10, n. 3, p. 273–296, 2005. 3



## ANEXO A – Tabelas

Tabela 7 – Análise de sensibilidade para 2014, com efeito resultado &gt; 1 e um efeito seleção &lt; 1 ( d &gt; 0 e s &lt; 0)

	p11	p10	p01	p00	Resultado	Seleção	Pr(U)	ATTND	Razão
<b>PSM (original)</b>	—	—	—	—	—	—	—	<b>-0.059</b>	<b>94.2%</b>
esc_superior	13%	2%	27%	3%	10.4	0.4	10.4%	-0.035	<b>96.6%</b>
esc_medio	44%	32%	44%	38%	1.2	0.8	38.4%	-0.056	<b>94.6%</b>
esc>=medio	57%	33%	70%	42%	3.3	0.6	48.4%	-0.036	<b>96.5%</b>
ocup_alto	4%	0%	9%	1%	11.4	0.4	3.2%	-0.051	<b>95.0%</b>
ocup_medio_sup	11%	2%	18%	4%	5.2	0.5	8.0%	-0.047	<b>95.4%</b>
ocup>=medio_sup	15%	2%	27%	5%	7.1	0.5	11.2%	-0.038	<b>96.3%</b>
formal	83%	70%	84%	75%	1.8	0.8	76.9%	-0.056	<b>94.6%</b>
ocup_pais_alto	2%	0%	5%	1%	5.6	0.3	1.8%	-0.054	<b>94.7%</b>
pais_esc_superior	6%	3%	14%	5%	2.8	0.4	6.6%	-0.052	<b>94.9%</b>
pais_esc_medio	21%	14%	22%	15%	1.6	0.9	17.5%	-0.056	<b>94.6%</b>

Tabela 8 – Análise de sensibilidade para 2014, com efeito resultado &lt; 1 e um efeito seleção &gt; 1 ( d &lt; 0 e s &gt; 0)

	p11	p10	p01	p00	Resultado	Seleção	Pr(U)	ATTND	Razão
<b>PSM (original)</b>	—	—	—	—	—	—	—	<b>-0.059</b>	<b>94.2%</b>
esc_sem_instrução	4%	10%	2%	6%	0.318	1.954	6%	-0.053	<b>94.8%</b>
esc_fund_incompleto	22%	36%	15%	32%	0.368	1.39	27%	-0.049	<b>95.2%</b>
esc_<=fund_incomp.	26%	46%	17%	38%	0.331	1.568	34%	-0.042	<b>95.9%</b>
ocup_baixo_inferior	2%	8%	1%	6%	0.228	1.307	5%	-0.059	<b>94.3%</b>
ocup_baixo_superior	11%	34%	8%	28%	0.237	1.438	22%	-0.05	<b>95.1%</b>
reg_nordeste	22%	38%	10%	23%	0.365	2.313	25%	-0.036	<b>96.5%</b>
pais_ocup_baixo_inf.	23%	37%	18%	34%	0.432	1.24	29%	-0.054	<b>94.7%</b>
pais_ocup_baixo_sup	24%	26%	18%	23%	0.755	1.324	23%	-0.054	<b>94.7%</b>
pais_esc_sem_instr.	19%	29%	16%	25%	0.576	1.27	23%	-0.055	<b>94.6%</b>
pais_esc_fund_incomp	37%	40%	35%	40%	0.791	1.044	38%	-0.058	<b>94.4%</b>
pais_esc<=fund_inc	7%	6%	8%	8%	0.996	0.838	7%	-0.059	<b>94.3%</b>

Tabela 9 – Interações dos grupos sociais do pai e do indivíduo

Código	Grupo ocupacional do pai	Grupo ocupacional do filho
ocup1_12	“Alto”	“Alto” e “Médio sup.”
ocup1_456	“Alto”	“Médio inf.”, “Baixo sup.” e “Baixo inf.”
ocup2_1	“Médio sup.”	“Alto”
ocup2_23	“Médio sup.”	“Médio sup.” e “Médio”
ocup2_456	“Médio sup.”	“Médio inf.”, “Baixo sup.” e “Baixo inf.”
ocup3_12	“Médio”	“Alto” e “Médio sup.”
ocup3_34	“Médio”	“Médio” e “Médio inf.”
ocup3_56	“Médio”	“Baixo sup.” e “Baixo inf.”
ocup4_12	“Médio inf.”	“Alto” e “Médio sup.”
ocup4_34	“Médio inf.”	“Médio” e “Médio inf.”
ocup4_56	“Médio inf.”	“Baixo sup.” e “Baixo inf.”
ocup5_12	“Baixo sup.”	“Alto” e “Médio sup.”
ocup5_34	“Baixo sup.”	“Médio” e “Médio inf.”
ocup5_56	“Baixo sup.”	“Baixo sup.” e “Baixo inf.”
ocup6_6	“Baixo inf.”	“Baixo inf.”
ocup6_45	“Baixo inf.”	“Médio inf.” e “Baixo sup.”
ocup6_123	“Baixo inf.”	“Alto”, “Médio sup.” e “Médio inf.”