

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS**

CARLA FEIJÓ COLOSSI

**DIFERENÇAS SALARIAIS POR GÊNERO NO MERCADO DE TRABALHO DA
REGIÃO METROPOLITANA DE PORTO ALEGRE**

Porto Alegre

2018

CARLA FEIJÓ COLOSSI

**DIFERENÇAS SALARIAIS POR GÊNERO NO MERCADO DE TRABALHO DA
REGIÃO METROPOLITANA DE PORTO ALEGRE**

Trabalho de conclusão de curso submetido ao Curso de Graduação em Ciências Econômicas da UFRGS, como requisito parcial da obtenção do título de Bacharel em Ciências Econômicas.

Orientador: Prof. Dr. Sérgio Marley Modesto Monteiro

Porto Alegre

2018

CIP - Catalogação na Publicação

Colossi, Carla Feijó
DIFERENÇAS SALARIAIS POR GÊNERO NO MERCADO DE
TRABALHO DA REGIÃO METROPOLITANA DE PORTO ALEGRE /
Carla Feijó Colossi. -- 2018.
75 f.
Orientador: Sérgio Marley Modesto Monteiro.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Faculdade
de Ciências Econômicas, Curso de Ciências Econômicas,
Porto Alegre, BR-RS, 2018.

1. Diferença salarial. 2. Discriminação de gênero.
3. Modelo Blinder-Oaxaca. I. Monteiro, Sérgio Marley
Modesto, orient. II. Título.

CARLA FEIJÓ COLOSSI

**DIFERENÇAS SALARIAIS POR GÊNERO NO MERCADO DE TRABALHO DA
REGIÃO METROPOLITANA DE PORTO ALEGRE**

Trabalho de conclusão de curso submetido ao Curso de Graduação em Ciências Econômicas da UFRGS, como requisito parcial da obtenção do título de Bacharel em Ciências Econômicas.

Orientador: Prof. Dr. Sérgio Marley Modesto Monteiro

**Porto Alegre
2018**

CARLA FEIJÓ COLOSSI

**DIFERENÇAS SALARIAS POR GÊNERO NO MERCADO DE TRABALHO DA
REGIÃO METROPOLITANA DE PORTO ALEGRE**

Trabalho de conclusão de curso submetido ao Curso de Graduação em Ciências Econômicas da UFRGS, como requisito parcial da obtenção do título de Bacharel em Ciências Econômicas.

Orientador: Prof. Dr. Sérgio Marley Modesto Monteiro

Aprovada em: Porto Alegre, ___ de _____ de 2018.

BANCA EXAMINADORA:

Prof. Dr. Sérgio Marley Modesto Monteiro – Orientador
UFRGS

Profa. Dra. Janice Dornelles de Castro
UFRGS

Prof. Dr. Sabino da Silva Porto Júnior
UFRGS

AGRADECIMENTOS

Gostaria de iniciar os agradecimentos com os professores da Faculdade de Ciências Econômicas, sem os quais esse trabalho não seria possível. Em especial, gostaria de agradecer ao Prof. Dr. Sérgio Marley Modesto Monteiro pelos ensinamentos, atenção e paciência ao longo da elaboração desse trabalho.

Também gostaria de agradecer aos meus colegas e amigos de faculdade por fazerem parte dessa jornada e contribuírem com debates e principalmente risadas e companhia. À Gabriele, Nathalia e Rochelle, já formadas, e à Ana Paula, Victoria, Luciano, Nicole e Fernanda. Um agradecimento especial à Daiane e ao Marcelo pela ajuda durante o trabalho.

Também devo agradecer à minha família. Aos meus pais, sem os quais eu não estaria aqui, e à Amora e ao Sardinha pela companhia nas longas noites de estudo.

Por fim e não menos importante, gostaria de agradecer à família Patas Dadas pelas oportunidades e ensinamentos que nunca imaginei que teria.

RESUMO

A diferença salarial entre homens e mulheres é um fato amplamente conhecido. Ainda assim, seus motivos são diversos, e análises diferentes apontam para diferentes explicações. O ponto de maior debate é em relação à presença ou não de discriminação na determinação da diferença salarial. Para melhor compreender o mesmo, o presente trabalho analisa as principais teorias econômicas de discriminação. Em especial, é exposta a teoria do gosto por discriminação, desenvolvida por Becker (1971), que é utilizada de base para o desenvolvimento de diversos modelos de decomposição da diferença salarial. O modelo de Blinder-Oaxaca (1973), assim como outros que utilizam a teoria do capital humano, estimativas de produtividade e foco nas relações interindustriais são utilizados para medir esse hiato salarial. Utilizando dados da região metropolitana de Porto Alegre, é realizado um estudo empírico sobre a diferença salarial entre gêneros. A diferença encontrada é de R\$ 0,60 por hora, e as evidências apontam para a presença de discriminação.

Palavras-chave: Diferença salarial. Discriminação de gênero. Modelo Blinder-Oaxaca.

ABSTRACT

The wage gap between men and women is a widely known fact. Still, their motives are diverse, and different analyzes point to different explanations. The most debated point is whether there discrimination or not in the wage difference. To better understand this, the present paper analyzes the main economic theories of discrimination. In particular, the theory of taste for discrimination, developed by Becker (1971), is presented, which is used as the basis for the development of several models of wage differential. The Blinder-Oaxaca (1973) model, as well as others that use human capital theory, productivity estimates and focus on interindustrial relations are used to measure this wage gap. Using data from the metropolitan region of Porto Alegre, an empirical study of the gender wage gap is made. The difference found is R\$ 0,60 per hour, and the evidence points to the presence of discrimination.

Keywords: Wage gap. Gender discrimination. Blinder-Oaxaca model.

LISTA DE ABREVIATURAS E SIGLAS

C010	– Sexo
C040	– Posição na Família
C050	– Cor
C060	– Estado Onde Nasceu
C102	– Última Série Concluída (Grau)
CBO	– Classificação Brasileira de Ocupações
CNAE	– Classificação Nacional de Atividade Econômica
DC	– Coeficiente de discriminação (<i>discrimination coefficient</i>)
FEE	– Fundação de Economia e Estatística
MDC	– Coeficiente de discriminação do mercado (<i>market discrimination coefficient</i>)
MQO	– Mínimos Quadrados Ordinários
PED	– Pesquisa de Emprego e Desemprego na Região Metropolitana de Porto Alegre
PNAD	– Pesquisa por Amostra de Domicílio
POS	– Posição na Ocupação dos Ocupados no Trabalho Principal (Padronizada)
Q422	– Quanto o Sr.(a) Ganhou pelo Trabalho Realizado no Mês Passado? (Remuneração Líquida no Trabalho Principal)
Q424	– Quanto o Sr.(a) Ganhou pelo Trabalho Realizado no Mês Passado? (Remuneração Líquida no Trabalho Adicional)
Q431	– Quantas Horas o Sr.(a) Trabalhou Efetivamente em Cada um dos Seus Trabalhos na Semana Passada (No Trabalho Principal)
Q432	– Quantas Horas o Sr.(a) Trabalhou Efetivamente em Cada um dos Seus Trabalhos na Semana Passada (No Trabalho Adicional)
Q40A	– Tempo de Emprego (Anos)
SIT	– Situação Ocupacional

SUMÁRIO

1	INTRODUÇÃO	9
2	TEORIA ECONÔMICA DA DISCRIMINAÇÃO	11
2.1	GOSTO POR DISCRIMINAÇÃO	11
2.2	DISCRIMINAÇÃO ESTATÍSTICA.....	19
2.3	CONSIDERAÇÕES FINAIS.....	25
3	MODELOS DE ESTIMAÇÃO DA DISCRIMINAÇÃO DE RENDIMENTOS POR GÊNERO	27
3.1	MINCER E POLACHEK E A TEORIA DO CAPITAL HUMANO	27
3.2	DECOMPOSIÇÃO DE BLINDER-OAXACA	30
3.2.1	Utilizando Mínimos Quadrados Ordinários	31
3.2.2	Utilizando Modelo De Heckman Em Dois Estágio	38
3.3	ESTIMAÇÃO DE PRODUTIVIDADE DE COX E NYE	42
3.4	MODELO DE DISCRIMINAÇÃO INTERINDUSTRIAL	43
4	ESTIMAÇÃO EMPÍRICA: DIFERENÇAS EM RENDIMENTOS POR GÊNERO NA REGIÃO METROPOLITANA DE PORTO ALEGRE	48
4.1	A BASE DE DADOS 48	
4.2	ANÁLISE DESCRITIVA DOS DADOS	49
4.3	AS VARIÁVEIS	57
4.4	MODELAGEM E ANÁLISE DOS RESULTADOS.....	61
5	CONSIDERAÇÕES FINAIS	71
	REFERÊNCIAS	73

1 INTRODUÇÃO

A diferença de rendimentos entre homens e mulheres já é um fato amplamente conhecido. Entretanto, suas razões seguem como motivo de debate entre economistas. Enquanto alguns argumentam que essa diferença se dá devido à discriminação, outros apontam para diferenças em produtividade entre os sexos. Para estudar essa diferença de rendimentos, é necessário primeiramente distinguir entre os diferentes tipos de discriminação. Claramente, também é necessário analisar quais variáveis influem nos salários e de que maneiras para possibilitar a formulação de modelos.

De acordo com dados da Organização das Nações Unidas - ONU (2017), em uma média global as mulheres recebem US\$ 0,77 para cada US\$ recebido pelos homens. Como na América Latina o salário representa a maior fonte de renda das famílias (MESSINA; SILVA, 2017), essa diferença se torna ainda mais significativa. Apesar da diminuição da diferença salarial entre gêneros na América Latina a partir dos anos 2000, a mesma ainda persiste (MESSINA; SILVA, 2017). Na região metropolitana de Porto Alegre, especificamente, esse processo de diminuição da diferença salarial foi interrompido com a recessão de 2016 e 2017, que atingiu mais fortemente as mulheres (FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA - FEE, 2018).

Com o intuito de entender melhor o diferencial de salários, diversos economistas formulam modelos econométricos de determinação de salários e medidas de discriminação. Dessa forma, se obtêm evidências mais robustas para argumentos que anteriormente eram embasados em estatísticas descritivas. O presente trabalho procura explicar os principais modelos desenvolvidos e seus resultados, porém não de forma exaustiva.

O entendimento da desigualdade de rendimentos entre gêneros proporciona um melhor entendimento da desigualdade de rendimentos como um todo, que tem uma papel importante na relação entre crescimento econômico e pobreza (MESSINA; SILVA, 2017), e a desigualdade de gênero é de grande importância pois mulheres são mais afetadas pela pobreza (ONU 2017). Além disso, o estudo da discriminação também é um teste para teorias padrão de economia (ARROW, 1998).

Além disso, conforme destacam Corcoran e Duncan (1979), o maior entendimento da diferença salarial faz com que se reavalie políticas que busquem

diminuir essa diferença e com que se compreenda melhor o funcionamento o mercado de trabalho. Os resultados dos autores, que serão explicitados em mais detalhes no capítulo 3, mostram que o mercado não trata igualmente trabalhadores igualmente produtivos. Desse modo, algumas hipóteses de perfeita competição em mercados em teorias econômicas necessitariam ser revisitadas. Cox e Nye (1989) apontam a importância de identificar as fontes de discriminação para a formulação de políticas eficazes.

Como a pesquisa considera apenas a classificação binária de sexo, gênero e sexo serão tratados como sinônimos durante o trabalho, apesar do reconhecimento da existência de uma extensa literatura diferenciando os dois.

O primeiro capítulo apresenta a teoria econômica da discriminação, que fornece a base teórica para o estudo. São apresentadas a teoria do “gosto por discriminação” de Becker (1971) e discriminação estatística. O segundo capítulo trata dos diferentes modelos econométricos de cálculo da diferença salarial e seus resultados. No terceiro capítulo, o modelo de Blinder-Oaxaca é utilizado para realizar um teste empírico, utilizando dados atuais da região metropolitana de Porto Alegre.

2 TEORIA ECONÔMICA DA DISCRIMINAÇÃO

De acordo com Becker (1971), discriminação de mercado se refere à rendas recebidas por diferentes grupos. Diferente desse conceito é o de segregação de mercado, que se refere à distribuição de emprego, ignorando os salários. A segregação existe quando membros de um mesmo grupo são empregados juntos em uma proporção significativamente maior do que ocorreria se a distribuição dos membros dos grupos em empregos fosse aleatória. Ainda segundo o autor, a discriminação de mercado ocorre devido a gosto por discriminação enquanto a segregação de mercado acontece como resultado de diferenças nesses gostos. As duas podem ocorrer isoladamente ou em conjunto (BECKER, 1971).

A discriminação afeta a diferença salarial de diferentes maneiras. Conforme Burnette (2008), é fundamental a divisão entre discriminação pré-mercado e no mercado. A diferença salarial só seria resultado da discriminação caso a mesma fosse de mercado. Se a discriminação pré-mercado faz com que mulheres tenham uma produtividade menor e seu salário menor é apenas reflexo disso, então os salários não são do tipo *customary wages*, ou seja, definidos por atos discriminatórios, e as hipóteses de eficiência do mercado se mantêm.

Mincer e Polachek (1974) fazem uma distinção similar ao diferenciar os efeitos da discriminação direta e os da indireta. Discriminação direta ocorre quando salários diferentes são pagos por empregadores pela mesma unidade de capital humano possuído por pessoas diferentes. Os efeitos da discriminação indireta ocorrem quando a discriminação no mercado diminui o incentivo de um determinado grupo de investir em capital humano voltado para o mercado. Desse modo, o menor grau de capital humano das mulheres, por exemplo, pode ser parcialmente resultado de expectativas de discriminação futura (MINCER; POLACHEK, 1974).

O presente capítulo apresenta as diferentes teorias econômicas formuladas sobre discriminação, seus conceitos, modelos e análises.

2.1 GOSTO POR DISCRIMINAÇÃO

Gary Becker em seu livro *The Economics of Discrimination* (1971) desenvolve uma teoria de discriminação no mercado de trabalho, preenchendo assim um vazio que, de acordo com o mesmo, os economistas haviam deixado até então. A teoria é

desenvolvida utilizando uma estrutura microeconômica, baseada nos princípios da utilidade, de modo que a discriminação é vista como uma desutilidade causada pelo contato com um ou mais indivíduos pertencentes a certo grupo. Assim, se um indivíduo possui “gosto por discriminação” (do inglês *taste for discrimination*, tradução nossa) ele age como se estivesse disposto a pagar (diretamente ou indiretamente na forma de rendimento reduzido) para interagir com o grupo de pessoas às quais ele discrimina (BECKER, 1971).

O coeficiente de discriminação (DC, do inglês *discrimination coefficient*, tradução nossa) serve como uma ponte entre os custos monetários e os líquidos quando há discriminação, que é tratada como um custo não-monetário. Além disso, a introdução do coeficiente torna possível definir de modo equiparável o “gosto por discriminação” para diferentes grupos de agentes no mercado, como empregadores, empregados e consumidores. A magnitude do coeficiente varia de pessoa para pessoa e de região para região, dependendo de outras variáveis além de preferências, como a distância física e social dos diferentes grupos de pessoas, seus status socioeconômicos, seu número relativo e sua substituíbilidade no processo produtivo (BECKER, 1971).

O autor segue definindo o coeficiente de discriminação do mercado (MDC, do inglês *market discrimination coefficient*). Assim como acontece com o coeficiente de discriminação, diversas variáveis além de preferências determinam o MDC, como, por exemplo, a distribuição completa dos DCs do mercado, a importância econômica dos grupos e o grau de competitividade do mercado, que é analisado adiante. O MDC seria igual à diferença da razão dos salários dos dois grupos no caso com e sem discriminação. Seguindo as notações utilizadas por Becker (1971) W representa o grupo que discrimina e N o grupo que é discriminado, π o salário e π^0 o salário quando não há discriminação, com os subscritos w e n identificando o salário de cada grupo.

$$MDC = \frac{\pi_w}{\pi_n} - \frac{\pi_w^0}{\pi_n^0}$$

Quando os dois grupos W e N são substitutos perfeitos no processo de produção, π_w^0 e π_n^0 são iguais, já que no cenário sem discriminação seus rendimentos seriam os mesmos. Nesse caso, o coeficiente de discriminação do

mercado assume a forma $MDC = \frac{\pi_w - \pi_n}{\pi_n}$. Se os dois grupos forem substitutos imperfeitos, mesmo em uma situação sem discriminação seus salários podem variar, de modo que se assume durante toda a análise que W e N são substitutos perfeitos (BECKER, 1971). Oaxaca (1973) utiliza esse coeficiente e o divide pela razão dos salários na ausência de discriminação para definir o seu coeficiente de discriminação.

A condição necessária e suficiente para ocorrer discriminação efetiva é W ser uma “maioria econômica” em uma proporção maior que N é uma “maioria trabalhadora”. Sendo Y_i a renda agregada de W ou N e l_i a oferta total de trabalho de cada grupo, a condição é representada por $\frac{Y_w}{Y_n} > \frac{l_n}{l_w}$ (BECKER, 1971).

Com esses conceitos e definições é possível analisar a discriminação e seus efeitos em diferentes grupos e tipos de mercado. Primeiramente analisa-se a discriminação por parte do empregador, e como ela se manifesta em mercados com diferentes graus de competitividade. Em seguida é estudado o caso da discriminação entre empregados, da discriminação por parte dos consumidores e a relação entre o coeficiente de discriminação e o governo.

O empregador discrimina ao se recusar a contratar um trabalhador cujo produto marginal excede o seu custo marginal. O custo não-monetário da discriminação é transformado em monetário utilizando o coeficiente de discriminação, representado por d . Assim, o salário com o qual o empregador se defronta é $\pi(1 + d)$. O princípio microeconômico da racionalidade se mantém, de modo que o empresário busca maximizar seu retorno líquido e minimizar seus custos líquidos (BECKER, 1971).

Para maximizar seu retorno líquido, o mesmo compara suas preferências com os salários de mercado. Considerando novamente dois grupos W e N e assumindo que o empregador discrimina contra N, ele compara o salário $\pi_n(1 + d)$ com π_w . Se $\pi_n(1 + d) > \pi_w$, apenas trabalhadores pertencentes ao grupo W são contratados; se $\pi_n(1 + d) < \pi_w$, apenas trabalhadores pertencentes ao grupo N são contratados; e se $\pi_n(1 + d) = \pi_w$, trabalhadores de ambos os grupos são contratados. Se $MDC = d$, então $\pi_n(1 + d) = \pi_w$, pois, como os grupos são substitutos perfeitos no processo de produção, $d = \frac{\pi_w - \pi_n}{\pi_n}$. Se for o caso, a lógica anterior se mantém: empregadores com coeficientes de discriminação menores que d empregariam apenas

empregados do grupo N e empregadores com seus DC maiores que d empregariam apenas membros do grupo W (BECKER, 1971).

A combinação de fatores que minimiza os gastos em um cenário com discriminação difere do cenário sem discriminação, pois há uma menor demanda para fatores que sofrem discriminação e pois os custos monetários são maiores que os mínimos. Os empregadores ainda igualam a razão dos produtos marginais (MP_i , do inglês *marginal product*, tradução nossa) com a razão dos salários (BECKER, 1971). Desse modo,

$$\frac{MP_i}{MP_j} = \frac{\pi_i(1 + d^i)}{\pi_j(1 + d^j)}$$

Um raciocínio similar é utilizado no caso de um mercado cujas firmas possuem funções de produção homogêneas. Quando isso ocorre, aquelas com maiores coeficientes de discriminação possuem um custo unitário maior e aquelas com DCs menores possuem custos menores. Conseqüentemente, a empresa com menor d venderia a um preço menor do que as demais e produziria o total vendido para o mercado. O MDC de equilíbrio seria igual ao DC dessa firma. Se as empresas não possuem funções de produção homogêneas, então o custo aumenta conforme aumenta a produção e a firma com menor coeficiente de discriminação não produz o total vendido para o mercado. Porém, no geral, firmas com DCs menores do que o MDC são mais lucrativas, por possuírem custos menores, e tendem a se expandir em relação às demais (BECKER, 1971).

Assim, no caso de mercados competitivos, o MDC de equilíbrio seria baixo, pois as empresas com DCs grandes sairiam do mercado por possuir custos maiores e apenas aquelas com DCs baixo permaneceriam. O mesmo ocorre no caso de monopólios transferíveis. Nessa situação, o monopolista com alto DC transferiria para um com DC menor, pois este faria a maior oferta já que, por ter um custo menor, teria o maior lucro. Se o monopólio não for transferível, a aleatoriedade assumida para a posse do monopólio faz com que coeficiente de discriminação de equilíbrio seja igual ao DC médio da classe de empregadores. Sendo assim, a discriminação é menor em mercados competitivos (BECKER, 1971).

Um exemplo empírico desse fenômeno é o caso das indústria de tabaco sueca do século XIX. As firmas que contratavam mais mulheres tinham maiores chances de sobreviver, pois o salario das mesmas era menor que o dos homens, e a

competição no mercado impedia as firmas de empregar homens a salários mais altos. Como homens possuíam outras oportunidades de trabalho com salários maiores, a força de trabalho dessa indústria se tornou predominantemente feminina (STANFORS et al., 2014).

Além da discriminação por parte do empregador, também é possível que ocorra discriminação entre os empregados. Quando isso ocorre, o empregado i que discrimina contra o grupo N tem seu rendimento líquido reduzido à $\pi_{in}(1 - d)$ ao trabalhar com outros empregados pertencendo ao grupo N . A discriminação de mercado dos empregados depende do gosto médio dos mesmos por discriminação e da dispersão em volta da média, o quão substituíveis os grupos de empregados são no processo de produção e da presença de sindicatos (BECKER, 1971).

O custo unitário c de discriminar pode ser calculado pela diferença entre trabalhar com o grupo no qual discrimina ou trabalhar com seu próprio grupo.

$$c = \frac{\pi_{in} - \pi_{iw}}{\pi_{in}}$$

Se o custo unitário de discriminar for maior do que o coeficiente de discriminação do empregado, ou seja, se $c > d$, o empregado trabalha com N . Se $c < d$ ele trabalha com W e se $c = d$ ele trabalha com ambos os grupos (BECKER, 1971).

No caso de W e N serem substitutos perfeitos no processo produtivo, ocorre segregação do mercado ao invés de discriminação ocorre, pois, se os empregados W possuem um gosto por discriminação contra N , eles exigem um salário maior para trabalhar com N do que para trabalhar com outros W . Um empregador, para minimizar seus custos, contrataria apenas empregados do grupo com menor salário, e W e N não são empregados pela mesma firma, ou seja, há segregação de mercado (BECKER, 1971).

Para analisar o caso de grupos complementares no processo produtivo se adiciona um terceiro grupo cujo trabalho é usado em proporções fixas com W e N e cujo coeficiente de discriminação contra N é dado por d_3 . Assim como os demais grupos este busca maximizar seu rendimento líquido, de modo que trabalha apenas com W caso seu salário líquido ao trabalhar com o mesmo seja maior do que o que receberia trabalhando com N , ou seja $\pi_{3w} > \pi_{3n}(1 - d_3)$; trabalha apenas com N se

$\pi_{3w} < \pi_{3n}(1 - d_3)$; e trabalha com ambos caso $\pi_{3w} = \pi_{3n}(1 - d_3)$. Considerando que são empregadas m unidades de trabalho do terceiro grupo para cada m' unidades de trabalho de W ou N, o custo unitário de produção é dado por $m\pi_{3i} + m\pi_i$, sendo o subscrito i igual a n ou w , dependendo do grupo que é empregado. Ambos os grupos N e W serão empregados apenas se $m\pi_{3n} + m\pi_n = m\pi_{3w} + m\pi_w$ (BECKER, 1971).

Discriminação entre empregados que são substitutos imperfeitos leva a um resultado que situa-se entre aquele quando os trabalhadores são substitutos perfeitos e quando são complementares (BECKER, 1971).

Toda a análise feita até agora sobre discriminação entre empregados supõe a inexistência de sindicatos, ou seja, em um mercado de trabalho competitivo. Quando os mesmos são introduzidos, os resultados mudam. Primeiramente, é possível que ocorra discriminação, não apenas segregação, entre trabalhadores substitutos perfeitos, se um grupo sindicalizado discrimina contra um grupo impedindo que eles entrem no sindicato. Becker (1971) realça o fato que diferenças salariais entre trabalhadores sindicalizados e não-sindicalizados podem surgir devido à discriminação contra outros grupos, não como consequência de discriminação do sindicato. Por exemplo, um grupo de homens ou de brancos pode ter um sindicato mais forte por terem sido os primeiros a entrar nessa profissão ou por serem mais ativos, porém parte da sua vantagem econômica pode se dar devido ao seu gênero ou cor, não por causa do sindicato (BECKER, 1971).

Sindicatos afetam o nível de emprego e a atratividade das ocupações. Um sindicato que impede a entrada de um grupo N aumenta o custo de contratar integrantes desse grupo, de modo que o salário líquido de N é representado por $\pi_n(1 + d)$. Se a presença sindical não altera a posição da curva de demanda, ela altera o ponto de um equilíbrio para um preço maior e uma quantidade menor, ou seja, menos pessoas são empregadas com um salário menor (BECKER, 1971).

As medidas que cada sindicato pode adotar dependem do seu poder. Um sindicato forte, ou seja, um que não enfrenta muita competição de outros trabalhadores ou de máquinas, consegue rejeitar muitos candidatos, já que o salário dos seus membros é inversamente relacionado com o número de associados. O contrário ocorre com um sindicato fraco, ou seja, um que enfrenta severa competição com outros trabalhadores ou máquinas: não é possível rejeitar um alto

número de candidatos pois o salário dos membros está diretamente associado com a quantidade de membros (BECKER, 1971).

A discriminação na escolha de membros não custaria nada aos já associados se o número de escolhidos de modo discriminatório fosse igual ao número de escolhidos ao acaso. Dessa forma, é de se esperar que mais discriminação ocorra em sindicatos fortes, já que os mesmos rejeitam um número maior de candidatos, e se não há custo por discriminar é de se esperar que isso ocorra (BECKER, 1971).

No entanto, outra maneira de controlar a entrada de membros é cobrando uma taxa de entrada, que deve ser igual à diferença entre o valor presente recebido por um membro e o valor presente do que seria recebido pela ocupação mais próxima que pague melhor. Quanto maior a taxa de entrada, maior o poder do sindicato. Nesse caso, não é possível discriminar sem custos, e o número de candidatos rejeitados seria menor (BECKER, 1971).

Em seguida é analisada a discriminação por parte dos consumidores. Conforme explicitado anteriormente, se dois grupos W e N são substitutos perfeitos no processo produtivo e não há discriminação, cada trabalhador recebe o equivalente ao seu produto marginal. No entanto, discriminação de consumidores contra N reduz o salário dos integrantes do mesmo em relação aos de W. Se os consumidores possuem coeficientes de discriminação diferentes, um aumento no número de N aumenta o coeficiente de discriminação de mercado, pois aqueles com maiores DCs são induzidos a comprar de N. A discriminação por parte dos consumidores aumenta o preço monetário P_n de um bem para um preço líquido $P_n(1 + d_c)$ para bens produzidos por N (BECKER, 1971).

Por fim, a discriminação por parte do governo é analisada. O governo pode discriminar diretamente algum grupo N aumentando os custos não monetários de empregadores contratarem os mesmos, por exemplo, restringindo a quantidade de horas ou o tipo de trabalho que esse grupo pode realizar. Isso aumenta o salário líquido de N do mesmo modo que a discriminação por parte do empregador, ou seja, $\pi_n(1 + d)$ (BECKER, 1971).

A discriminação do governo também tem consequências políticas. Assume-se que nas eleições o único assunto a ser votado é a política em relação a dois grupos e que as preferências de cada eleitor são dadas por seu coeficiente de discriminação. Como cada indivíduo votará no partido cujo DC é mais próximo do seu, o DC mediano é o único equilíbrio possível. Uma aplicação empírica desse

modelo não traz resultados significativos pois, como aponta o autor, para o mesmo são necessárias as suposições que a política afeta apenas os eleitores, não toda a população, que as preferências individuais em relação ao comportamento do governo diferem das preferências pessoais e, possivelmente a mais relevante, que cada eleição só decide um problema, enquanto no mundo real um voto expressa opiniões em relação a diversas questões, não apenas discriminação. Entretanto, a existência de diversas pautas em uma mesma eleição é uma das explicações da grande importância dos votos das minorias, pois, por exemplo, a posição dos candidatos pode dividir os votos do grupo W na metade, porém suas diferentes políticas discriminatórias podem fazer com que N vote em um deles e assim decida a eleição (BECKER, 1971).

A soma da discriminação causada por cada grupo descrito acima (empregadores, empregados, consumidores e governo) resulta na discriminação de mercado efetiva. O salário líquido de N é então descrito como $\pi_n(1 + d_a + d_b + d_c)$, onde d_a representa o coeficiente de discriminação dos empregadores, d_b o do governo e d_c o dos sindicatos (BECKER, 1971).

Uma das críticas à teoria apresentada acima é que adicionar mais uma variável (no caso d) arriscaria transformar a explicação em tautologia (ARROW, 1998). Ao criticar o estudo de Becker, Arrow (1998) afirma que, se os gostos não forem homogêneos, os empregadores que menos discriminam eliminariam os demais do mercado, e se houvessem alguns que não discriminam não haveria discriminação. Porém, conforme exposto acima, Becker (1971) faz essa mesma análise.

Outra crítica feita por Arrow (1998) é que boa parte das contratações são feitas por grandes empresas, então o autor estaria atribuindo gostos a entidades impessoais. Além disso, não seria do interesse dos investidores que a empresa discriminasse (ARROW, 1998). Apesar de grandes empresas não serem pessoas físicas, normalmente aqueles que realizam a contratação são, por exemplo os departamentos de relações humanas, de modo que os gostos seriam atribuídos aos mesmos.

Por fim, o autor aponta que a teoria do gosto por discriminação explica a segregação entre firmas (conforme apresentado anteriormente), porém não a segregação entre ocupações (ARROW, 1998). Becker (1971) elabora brevemente sobre o assunto, afirmando que mesmo a proporção de N em cada ocupação

variando, se todos os demais grupos trabalhando com N tivessem o mesmo DC, a taxa de salário relativo de N seria igual em todas. Ele também afirma que os membros do grupo discriminado tenderiam a trabalhar mais com membros do mesmo grupo na mesma ocupação para evitar discriminação (BECKER, 1971).

2.2 DISCRIMINAÇÃO ESTATÍSTICA

Discriminação de um grupo no mercado de trabalho ocorre quando o salário médio do mesmo não é proporcional à sua produtividade média (AIGNER; CAIN, 1977). A discriminação estatística (do inglês *statistical discrimination*, tradução nossa) ocorre quando a informação é de custosa obtenção, fazendo com que os empregadores baseiem suas decisões de empregar ou pagar os trabalhadores nas características do grupo ao qual o mesmo pertence (CORCORAN; DUNCAN, 1979). Assim, a raça ou o sexo dos trabalhadores são usados como *proxy* de outras características não-observáveis (PHELPS, 1972).

Arrow (1998) critica essa teoria ao afirmar que, se existem um número de variáveis observáveis, como escolaridade, então a teoria implicaria que adicionar raça ou gênero como variáveis para estimar os salários melhoraria a previsão. Porém o mesmo apenas ocorreria caso todas as variáveis fossem observáveis já que, como afirmou Phelps (1972), raça e gênero são apenas utilizadas como *proxy* de variáveis não-observáveis.

Dickinson e Oaxaca (2009) fazem uma distinção entre dois tipos de discriminação estatística. A primeira, que é a considerada na maioria dos estudos, ocorre quando os empregadores baseiam suas decisões na produtividade média do grupo ao qual o indivíduo pertence (no caso analisado, mulheres). O segundo tipo ocorre quando as decisões são baseadas na percepção de risco de agentes avessos ao risco. O risco pode ser entendido pelos empregadores em relação à variância da produtividade, à sua distribuição ou em relação à probabilidade de perda. O experimento feito pelos autores mostra grande significância para esse segundo tipo de discriminação estatística, principalmente quando a probabilidade de perda é grande (DICKINSON; OAXACA, 2009).

Phelps (1972) desenvolve um modelo básico para analisar a discriminação estatística. É assumido que os empregadores utilizam a nota y_i de um teste para medir a produtividade ou nível de qualificação q_i dos candidatos. Desse modo,

$y_i = q_i + \mu_i$, onde μ_i é o termo de erro normalmente distribuído e com média zero (PHELPS, 1972). Como os empregadores estão interessados em y_i apenas na medida em que o mesmo da informações sobre a variável não observada q_i , o verdadeiro interesse dos empregadores é o valor esperado de q_i , ou seja, \hat{q}_i . Sendo α a média do grupo e γ o nível de confiabilidade do teste y , onde $0 < \gamma < 1$. Conforme os empregadores assimilem mais informações com o passar do tempo, γ tenderia a aumentar. Em mercados competitivos a média é não viesada, pois a expectativa dos empregadores em relação a q é correspondente à expectativa realizada, já que a competição faz com que os erros não se repitam (AIGNER; CAIN, 1977). Assim,

$$\hat{q}_i = E(q|y) = (1 - \gamma)\alpha + \gamma y$$

Supondo que os empregadores também podem observar outra característica do grupo, como raça ou gênero (o autor utiliza raça durante toda sua análise, porém ressalta que a análise de discriminação de gênero é análoga). Nesse caso, a qualificação seria determinada por $q_i = \delta + x_i + \eta_i$, onde x_i seria a “contribuição dos fatores sociais” $x_i = (-\beta + \epsilon_i)c_i$, sendo $c_i = 1$ se o candidato for negro (ou mulher) e η_i e ϵ_i são os termos de erro normalmente distribuídos (PHELPS, 1972).

Como apontam Aigner e Cain (1977), x_i resulta em um valor médio menor para o q_i da minoria discriminada, pois o coeficiente negativo $-\beta$ faz com que, caso $c_i = 1$, a contribuição dos fatores sociais seja negativa. A justificativa para o mesmo é dada por Phelps (1972) no início de seu texto é ambígua (AIGNER; CAIN, 1977), e remete às experiências anteriores dos empregadores, fatores sociais desvantajosos ou à estereótipos (PHELPS, 1972). Aigner e Cain (1977) apontam que, em mercados competitivos, estereótipos não permanecem pois os empregadores não insistem em comportamentos errôneos, que fazem que os mesmos tenham maiores gastos ou percam lucro. Todavia, Arrow (1998) alega que cada empregador possui uma quantidade limitada de experiência, de forma que suas crenças anteriores podem permanecer intactas por certos períodos de tempo. Aigner e Cain (1977) também realçam que essa diferença na produtividade média assumida na partida é problemática, já que discriminação econômica é definida como diferença em pagamentos não relacionada com diferenças em habilidades (AIGNER; CAIN, 1977).

Sendo $\lambda_i = \eta_i + c_i \epsilon_i$ e $z_i = -\beta c_i$, é possível reescrever as equações de modo que $q_i = \delta + z_i + \lambda_i$ e $y_i = q_i + \mu_i = \delta + z_i + \lambda_i + \mu_i$. Desse modo, utilizando a estatística de teste se obtém um modelo de regressão único para ambas as raças (ou sexos), no qual o termo da direita é igual a zero se o indivíduo for branco (ou homem).

$$q'_i = \frac{\text{var}\lambda_i}{\text{var}\lambda_i + \text{var}\mu_i} y'_i + \frac{\text{var}\lambda_i}{\text{var}\lambda_i + \text{var}\mu_i} z'_i + u_i$$

Aigner e Cain (1977) apontam uma falha na regressão acima. Como Phelps (1972) supõe que a inclinação de y em q é diferente para os dois grupos analisados, ou seja, que $\text{var}(q^B) > \text{var}(q^W)$. Essa maior variância significa que os resultados dos testes são mais confiáveis para medir as habilidades de negros do que de brancos (PHELPS, 1972). No entanto, Aigner e Cain (1977) ressaltam que seria necessário um termo de interação yz na regressão única. Além disso, Phelps (1972) apresenta essa hipótese sem nenhuma evidência empírica, e argumentos podem ser feitos sobre maiores variâncias para os dois grupos, por exemplo, negros (mulheres) podem enfrentar mais restrições externas, fazendo sua variância ser mais homogênea, ou como brancos (homens) enfrentam condições externas determinantes de q similares, a variância destes poderia ser menor (AIGNER; CAIN, 1977).

Além disso, a suposição de diferentes inclinações implica que, para valores acima de um certo y , negros recebem mais do que brancos (PHELPS, 1972). Porém as evidências empíricas apontam o contrário: considerando y como anos de experiência ou de escolaridade, conforme o mesmo aumenta a diferença salarial (tanto entre raças quanto entre gêneros) também aumenta (AIGNER; CAIN, 1977).

Por fim, Aigner e Cain (1977) concluem que o modelo proposto por Phelps (1972) é inadequado para descrever discriminação econômica, pois cada trabalhador é pago de acordo com sua produtividade esperada.

Tendo em vista as limitações do modelo de Phelps (1972), Aigner e Cain (1977) propõem um modelo alternativo para discriminação estatística. O grande diferencial do mesmo é que os empregadores não são neutros em relação ao risco, como no modelo anterior. No modelo proposto, os empregadores buscam maximizar a utilidade da produtividade, $U(q)$, a , b e c são parâmetros da função utilidade e ϵ a base do logaritmo natural.

$$U(q) = a - b\epsilon^{-cq}$$

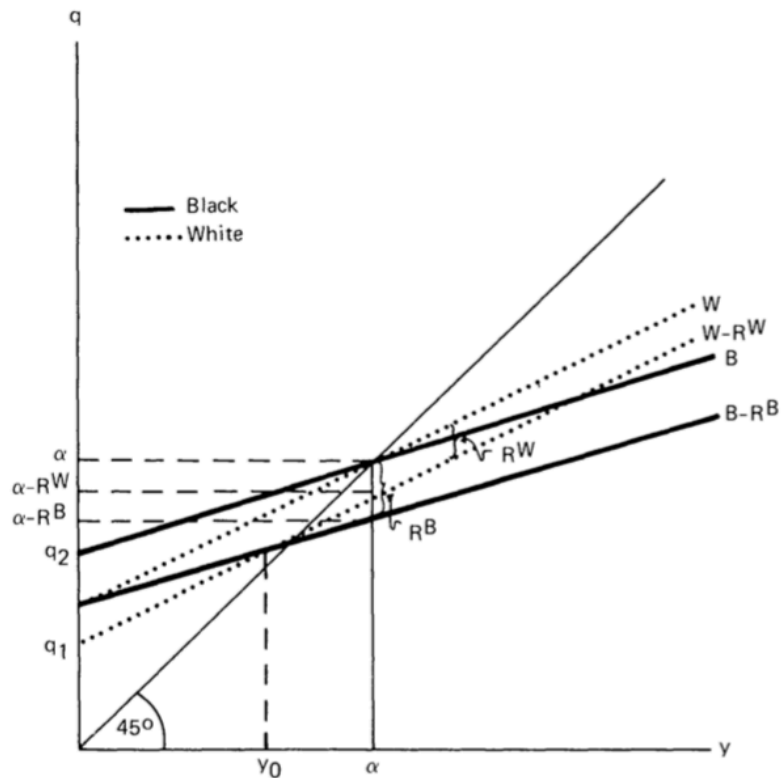
$$E[U(q|y)] = a - b\epsilon^{-cE(q|y) + c^2/2 \text{var}(q|y)}$$

Maximizar $E[U(q|y)]$ é equivalente à maximizar seu logaritmo, que é o mesmo que maximizar $E(q|y) + k\text{var}(q|y)$, onde $k = \frac{c^2}{2}$. Assim, o risco é definido como $k\text{var}(q|y)$, e é independente do nível de y . Dessa forma, o empregador irá escolher quem empregar de modo a maximizar o valor esperado de q , descontando o risco (AIGNER; CAIN, 1977). Conforme discutido anteriormente, as evidências empíricas encontradas por Dickinson e Oaxaca (2009) mostram grande significância do risco para a ocorrência de discriminação estatística.

A figura 1 mostra graficamente a relação entre q e y . As linhas W e B são dadas pela equação $\hat{q}_i = E(q|y) = (1 - \gamma)\alpha + \gamma y$, descrita anteriormente, assumindo $\alpha^W = \alpha^B$, ou seja, a produtividade média é igual para os dois grupos, e $\gamma^B < \gamma^W$, que significa que os resultados dos testes são mais confiáveis para brancos (W) do que para negros (N). As linhas W-R^W e B-R^B representam $E(q|y) + k\text{var}(q|y)$ para cada um dos grupos, ou seja, o valor esperado da produtividade descontado o risco. De acordo com os autores, a discriminação econômica é aparente pois N recebe menos para uma mesma produtividade esperada (AIGNER; CAIN, 1977).

Uma característica realista do modelo é que a diferença salarial aumenta conforme a variável que indica produtividade (y) aumenta. Como descrito anteriormente, estudos empíricos apontam para um maior diferencial de ganhos para níveis mais altos de experiência e escolaridade. No entanto, conforme pode ser visto na figura 1, o modelo aponta para rendimentos maiores para negros do que para brancos até certo nível y_0 , e os autores apontam que não há evidências empíricas que comprovem o mesmo. O mesmo só seria corrigido no caso de um alto fator de risco (no caso equivalente à $q_2 - q_1$). Aigner e Cain (1977) apontam alguns motivos para o fator de risco não ser tão grande, entre eles que as empresas grandes teriam capacidade de se assegurar contra riscos em relação à produção, em mercados de capitais perfeitos as pequenas empresas poderiam fazer o mesmo.

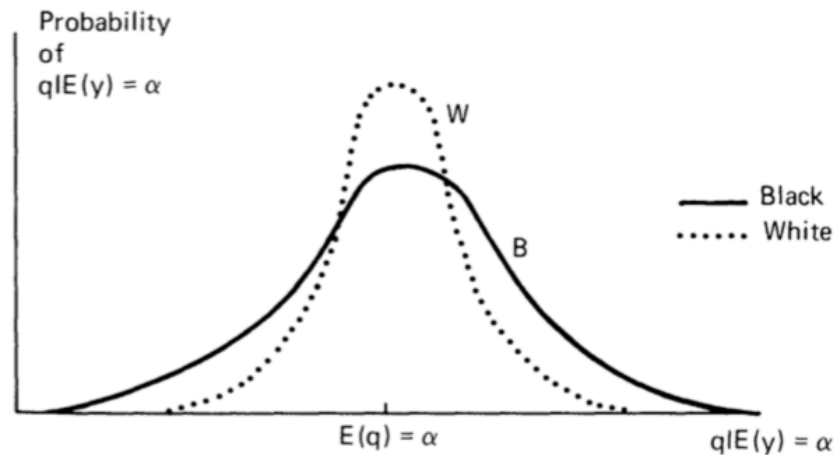
Figura 1 - Previsões de Produtividade (q) por Raça e Resultado de Teste (y) com Desconto de Risco



Fonte: Aigner e Cain (1977).

Sendo assim, o risco depende da variância condicional de q por y dos dois grupos, sendo $var(q^W) = var(q^B)$ onde os sobrescritos W e B representam a produtividade de brancos e negros, respectivamente (lembrando que o modelo também pode ser utilizado para discriminação de gênero). Assumindo uma variância condicional de q menor para brancos, os mesmos apresentariam um menor fator de risco. Posto isto, para um mesmo valor esperado de teste $y = \alpha = E(y)$, o menor valor recebido por negros (ou mulheres) se dá devido ao maior fator de risco (AIGNER; CAIN, 1977). A Figura 1 mostra graficamente a distribuição neste ponto $y = \alpha = E(y)$ e para $\alpha = E(q)$ para os dois grupos.

Figura 2 - Distribuição da Produtividade (q), Dado o Resultado de Teste $y = E(y)$



Fonte: Aigner e Cain (1977).

A discriminação, no caso descrito acima, seria resultado de instrumentos inadequados de teste, que teriam como consequência os diferentes fatores de risco, e não “gosto por discriminação” (AIGNER; CAIN, 1977).

Os autores não testam empiricamente o modelo. Além disso, se a definição de discriminação utilizada incorporar a aversão ao risco por parte dos empregadores, o modelo deixaria de representar discriminação econômica (AIGNER; CAIN, 1977).

Outro modelo de discriminação estatística é desenvolvido por Spence (1973¹ *apud* AIGNER; CAIN, 1977). Assim como nos demais, a produtividade q é uma variável não-observável, fazendo com que os empregadores baseiem suas decisões nos resultados de testes, y . Os trabalhadores podem investir tempo e dinheiro para aumentar y , e esse custo é inversamente relacionado com q . O sistema de salários leva a um equilíbrio onde q e y são diretamente relacionados, de modo que as expectativas dos empregadores se confirmam na realidade (AIGNER; CAIN, 1977).

Nesse modelo, o nível de y que separa trabalhadores com níveis baixos e altos de habilidade é arbitrário. Desse modo, esse nível y^* pode ser diferente para grupos diferentes (raciais ou de gênero) com mesma produtividade. Assim, se y^* for mais alto para negros ou mulheres do que para brancos ou homens mas ainda assim baixo o suficiente que os compense a obter esse nível y^* , então brancos ou homens terão um salário maior, pois seu custo para atingir seu y^* foi menor. Dessa

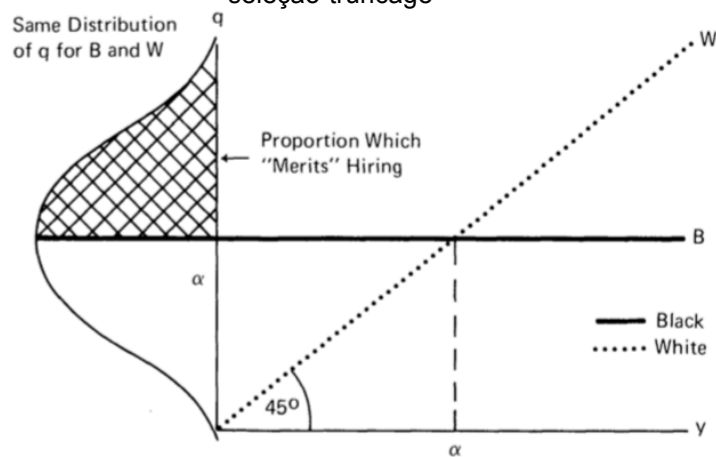
¹ SPENCE, Michael. Job Market Signaling. *The Quarterly Journal of Economics*. v. 87, n. 3, p. 355-374, Ago. 1973. Disponível em <<https://bit.ly/2KCzpFI>>. Acesso em 20 jul 2018.

forma, seria possível existir diferença salarial em mercados competitivos (AIGNER; CAIN, 1977).

No entanto, Aigner e Cain (1977) questionam a estabilidade desse modelo. Se cada empregado conhece sua própria produtividade ou seu custo para obter y^* , então aquele grupo sendo discriminado saberá que está sendo pago menos do que aqueles do outro grupo com produtividade ou custo de obtenção de y^* equivalente. Em equilíbrio haveria uma oferta salarial comum para mesmos níveis de y , e conseqüentemente não haveria discriminação econômica (AIGNER; CAIN, 1977).

Por fim, existe o modelo “seleção-truncagem” de discriminação estatística. Esse modelo supõe que as decisões de contratação estão confinadas na parte superior da distribuição de y , sendo trabalhadores com um nível de produtividade inferior a α não contratados. Como os resultados de testes são mais confiáveis para brancos (ou homens), ou seja, seu γ é maior, eles são preferidos com base em seus maiores valores esperados de q , mesmo a distribuição de q para ambos os grupos sendo igual (AIGNER; CAIN, 1977). A Figura 3 exemplifica um caso extremo, onde $\gamma^W = 1$ e $\gamma^B = 0$.

Figura 3 - Previsões de Produtividade (q) por Raça e Resultado de Teste (y) – Modelo “seleção-truncagem”



Fonte: Aigner e Cain (1977).

2.3 CONSIDERAÇÕES FINAIS

O presente capítulo apresentou os principais tipos de discriminação analisados pela teoria econômica. A teoria do “gosto por discriminação” desenvolvida por Becker (1971) analisa o coeficiente de discriminação, que é

utilizado juntamente com as demais variáveis nas determinações de relações de mercado.

Já a teoria da discriminação estatística possui diversas formas, sendo sua base a discriminação de grupos com base em sua característica comum (como gênero ou raça) devido à falta de informação. A presença de testes para determinar as habilidades dos indivíduos e o risco são incluídos em análises dessa teoria.

Como a segunda exige uma análise empírica de mercados e tipos de ocupação separadamente, pois, por exemplo, mulheres acabariam sendo empregadas predominantemente em ocupações de baixo rendimento devido à discriminação estatística, a primeira teoria é mais utilizada na formulação de modelos teóricos para análise de mercados como um todo. Ainda assim, durante algumas análises são observadas ocorrências de discriminação estatística.

Tendo essa base teórica, é possível a construção de modelos que buscam entender se a diferença salarial ocorre devido à discriminação. Sendo assim, o próximo capítulo examina os principais modelos de estimação de discriminação salarial, assim como apresenta seus resultados.

3 MODELOS DE ESTIMAÇÃO DA DISCRIMINAÇÃO DE RENDIMENTOS POR GÊNERO

A estimação da diferença salarial pode ser feita e analisada de diversas formas e com diversos enfoques. O presente capítulo busca fazer uma revisão dos principais modelos que buscam fazer essa estimação, assim como inspecioná-la para determinar o quanto da mesma se dá devido à discriminação, assim como apresentar os resultados encontrados. Com exceção do modelo desenvolvido por Cox e Nye (1989), todos os demais encontram evidências de discriminação de rendimentos por gênero. Como será analisado, este último se destaca por ser o único que busca medir a produtividade, porém vale ressaltar que também é o único que não utiliza dados atuais da época que foi escrito.

3.1 MINCER E POLACHEK E A TEORIA DO CAPITAL HUMANO

Ao analisar com *cross-section* os dados da participação das mulheres no mercado de trabalho nos países desenvolvidos a partir da década de 1950, fica claro o padrão de “dois picos”, ou seja, alta participação assim que entra no mercado de trabalho, seguida por uma queda que dura alguns anos (normalmente o período da primeira gravidez até a criança completar uma certa idade) seguido por uma volta ao mercado. Esses achados resultam na noção de carreiras interrompidas para mulheres (GOLDIN, 1983).

Mincer e Polachek (1974) desenvolvem um modelo que leva em consideração os períodos de abstinência das mulheres no mercado de trabalho. Utilizando a teoria do capital humano como base, seu objetivo é analisar como a acumulação de capital durante a vida das mulheres influem nos seus ganhos, e assim interpretar esses investimentos considerando expectativas passadas e futuras (MINCER; POLACHEK, 1974). De acordo com a teoria do capital humano, os ganhos dos indivíduos é uma função de seu capital humano acumulado.

$$E_t = E_{t-1} + rC_{t-1} - \delta_{t-1}E_{t-1}$$

Sendo E os ganhos no período t , C o total investido em capital humano no período $t - 1$, δ_{t-1} a taxa de depreciação do estoque de capital humano e r a taxa

de retorno médio do investimento em capital humano de um indivíduo. Como indivíduos recebem os retornos dos investimentos em capital humano relacionados ao trabalho ao realizar o próprio, a realização de tais investimentos depende do tempo em que cada indivíduo permanecerá no mercado. Mulheres passam em média menos tempo no mercado de trabalho devido às funções domésticas e de criação de filhos que lhes são atribuídas, além desse tempo ser descontínuo devido a saída para exercer essas funções. Devido a esses fatores, os autores concluem que mulheres casadas possuem uma chance menor de ter adquirido treinamento de habilidades e outros investimentos de capital humano para o trabalho. Além disso, expectativa de descontinuidade no trabalho pode fazer com que mulheres invistam menos (MINCER; POLACHEK, 1974).

Essa análise é consistente com os resultados encontrados por Oaxaca (1973), cujas regressões mostraram que para um mesmo coeficiente de retorno para o treinamento no trabalho, homens investem mais inicialmente e por um período de tempo maior. Por outro lado, a teoria mercado de trabalho interno sugere que parte do diferencial salarial de gênero entre indústrias ocorre devido às diferentes ligações dos gêneros com o trabalho. Dessa forma, as firmas podem considerar homens como membros mais permanentes no mercado de trabalho e conseqüentemente investir mais no seu treinamento e pagar mais para os mesmos (FIELDS; WOLFF, 1995).

Para enquadrar essa descontinuidade em um modelo, Mincer e Polachek (1974) dividem a história das mulheres no mercado (tradução livre do inglês *work histories*) entre períodos de participação (e) e não participação (h). Para os dados analisados foram considerados dois períodos de não participação, o primeiro antes do nascimento do primeiro filho e o segundo depois do nascimento do primeiro filho, podendo também existir um terceiro período no caso de mais filhos ou saídas por outros motivos. No caso de mulheres nunca casadas, o primeiro período de participação se estende durante todo seu tempo no mercado. Para simplificar o modelo, se assume investimentos constantes em cada segmento, apenas variando entre segmentos, onde o investimento k é expresso por $k_i = a_i + b_i t$, a_i a taxa de investimento no segmento i e b_i é a mudança na taxa de investimento por segmento. Essa formulação indica que pode haver depreciação mesmo durante períodos de participação no mercado (MINCER; POLACHEK, 1974).

$$\ln E_t = \ln E_o + \sum_i (rk_i - \delta_i)$$

Para estimar as taxas de salários das mulheres, é utilizada a função $\ln(w) = f(S, e, h, x) + u_i$, onde w representa o salário por hora, S os anos de estudo, e o vetor dos segmentos de participação no mercado, h o vetor dos segmentos de não participação no mercado e x o vetor que controla pelas outras variáveis (número de crianças, treinamento, horas de trabalho, mobilidade e saúde). Utilizando esse modelo são rodadas regressões utilizando dados do *National Longitudinal Survey of Work Experience* de 1967 (MINCER; POLACHEK, 1974).

Os resultados mostram que os coeficientes de investimento em capital humano aumentam de mulheres casadas com crianças para mulheres casadas sem crianças para mulheres solteiras. No entanto o grau de escolaridade não mostrou um padrão consistente em relação aos coeficiente de investimento da mesma forma que o estado civil. Os coeficientes de h foram bem baixos, porém negativos, representando uma depreciação no poder aquisitivo. Um resultado surpreendente é que a variável crianças não se mostrou estatisticamente significativa. A variável treinamento teve resultados diferentes para mulheres casadas e solteiras, mostrando um investimento menor em treinamento das mulheres que trabalham continuamente (ou seja, solteiras) e um investimento maior em mulheres cuja permanência no mercado é descontínua (ou seja, casadas). Mobilidade também apresentou resultados diferentes para mulheres com estados civis diferentes, pois a mudança de cidade tende a ser para melhorar o salário de mulheres solteiras (ou seja, uma mudança endógena) e tende a ser devido a realocação do marido para mulheres casadas (ou seja, exógena), de modo que o sinal do coeficiente varia para as duas (MINCER; POLACHEK, 1974).

Assim como Oaxaca, Mincer e Polachek (1974) rodam regressões separadas para mulheres negras já que a maioria dos estudos apontam que mulheres brancas e negras possuem funções de oferta de trabalho diferentes (GOLDIN, 1983). Os coeficientes de experiência para as mesmas tem metade do tamanho do que os das mulheres brancas, além do coeficiente de depreciação para tempo em casa não ser estatisticamente significativo. Isso implica menos investimento em capital humano no trabalho, apesar delas passarem mais tempo no mercado que mulheres brancas.

Utilizando esse modelo para explicar o diferencial salarial, os autores concluem que o mesmo explica 70% da diferença observada entre os salários de

homens e mulheres casados e 50% da diferença salarial entre homens casados e mulheres solteiras. Os autores não concluem que o restante seria necessariamente resultado da discriminação, explicando como na teoria do capital humano um menor ganho das mulheres representa uma medida do investimento na família no capital humano dos filhos. Não há discussão no texto dos motivos dessa função ser atribuída à mulher. O trabalho de cuidados domésticos e dos filhos é o típico caso da naturalização descrito por Saffioti (2001), quando funções atribuídas socialmente são normalizadas e tidas como naturais ou de justificativa biológica.

O modelo também apresenta outros problemas. Como os próprios atores apontam, há um problema de simultaneidade e de correlação entre experiência passada e ganhos passados e futuros, de modo que a variável experiência estaria sendo tanto determinada como um determinante da função de ganhos. Isso faria com que os resíduos fossem correlacionados com a variável experiência, tornando os estimadores viesados. Para tentar resolver esse problema, Mincer e Polachek (1974) estimam a função de ganhos utilizando mínimos quadrados em dois estágios e comparam os resultados com aqueles obtidos anteriormente utilizando mínimos quadrados ordinários. Como os resultados foram praticamente os mesmos, os autores relevam o problema (MINCER; POLACHEK, 1974).

No entanto, Duncan (1974) aponta outro possível problema com os resíduos do modelo: todos os fatores dados como além do escopo do modelo, como fatores discriminatórios institucionais e culturais, seriam capturados pelo erro, fazendo com que o mesmo fosse um fator causal comum dos ganhos e da experiência. Duncan (1974) também aponta outras dificuldades do modelo. Os autores não estimam uma função de fertilidade, que é justificado pelos mesmos no texto por envolver as mesmas variáveis que a função de oferta, que Duncan (1974) considera inadequada. Além disso, o autor critica a falta de tratamento das atividades dos outros membros da família, que são muito mencionados, e de uma medida ou *proxy* para a qualidade da experiência.

3.2 DECOMPOSIÇÃO DE BLINDER-OAXACA

Provavelmente o modelo mais utilizado para analisar a diferença salarial é o conhecido como decomposição de Blinder-Oaxaca. Ambos desenvolvem métodos similares no mesmo ano, 1973. Apesar de divergências em relação à abordagem,

ambos propõe uma divisão da diferença salarial entre diferença em características individuais dos sexos e diferença nos retornos dessas características, ou seja, discriminação.

3.2.1 Utilizando Mínimos Quadrados Ordinários

Apesar de similares, as abordagens de Oaxaca (1973) e Blinder (1973) diferem, de modo que serão apresentados separadamente, começando pelo segundo. Blinder (1973) inicia definindo a função de salários $Y_i = \beta_0 + \sum_{i=1}^n \beta_i X_{ji} + u_i$, onde Y_i é o logaritmo natural dos salário e X_{ji} são as n características observáveis de cada indivíduo i . Como o interesse é a comparação dos rendimentos de dois grupos, sendo o com sobescrito H aquele com a renda maior (no caso estudado, homens) e com sobescrito L aquele com renda menor (mulheres).

$$Y_i^H = \beta_0^H + \sum_{i=1}^n \beta_i^H X_{ij}^H + u_i^H$$

$$Y_i^L = \beta_0^L + \sum_{i=1}^n \beta_i^L X_{ij}^L + u_i^L$$

O diferencial salarial, então, seria igual à diferença das duas equações, que pode ser divididas entre a parte explicada pela regressão $\sum_{i=1}^n \beta_i^H X_{ij}^H - \sum_{i=1}^n \beta_i^L X_{ij}^L$ e a parte, normalmente atribuída à discriminação, capturada pelo intercepto, ou seja $\beta_0^H - \beta_0^L$, que seria a parte não explicada. No entanto, a parte explicada pela regressão também é dividida, entre a explicada pela diferença nos coeficiente e a explicada pela diferença nas características pessoais, de modo que pode ser reescrita.

$$\sum \beta_i^H X_{ij}^H - \sum \beta_i^L X_{ij}^L = \sum \beta_i^H (\bar{X}_i^H - \bar{X}_i^L) + \sum \bar{X}_i^L (\beta_i^H - \beta_i^L)$$

O primeiro termo do lado direito da igualdade é a diferença “atribuída às dotações”, ou seja, as vantagens em características individuais possuídas pelo grupo H, avaliadas pelo coeficiente da equação salarial do mesmo. O segundo termo seria o “atribuído aos coeficientes”, ou seja, a diferença entre como a equação

salarial do grupo H iria avaliar as características individuais dos mesmos e como elas são realmente avaliadas pela equação salarial do grupo L. O segundo termo só existe pois o mercado recompensa desigualmente os atributos dos membros dos diferentes grupos, de modo que a diferença “atribuída aos coeficientes” também se dá devido à discriminação (BLINDER, 1973).

Assim como é explicado adiante no modelo de Oaxaca (1973), onde os coeficientes utilizados podem ser tanto a estrutura salarial dos homens quanto das mulheres, também seria possível separar o termo $\sum_{i=1}^n \beta_i^H X_{ij}^H - \sum_{i=1}^n \beta_i^L X_{ij}^L$, colocando em evidência β_i^L , porém a interpretação econômica é mais direta utilizando a formulação explicitada anteriormente.

Resumidamente, tem-se R como diferencial bruto, E como a porção do diferencial “atribuída às dotações”, C como a porção do diferencial “atribuída aos coeficientes”, U como a porção não explicada do diferencial e D como a porção do diferencial atribuída à discriminação.

$$R = \beta_0^H + \sum_{i=1}^n \beta_i^H X_{ij}^H - \left(\beta_0^L + \sum_{i=1}^n \beta_i^L X_{ij}^L \right) = E + C + U$$

$$E = \sum \beta_i^H (\bar{X}_i^H - \bar{X}_i^L)$$

$$C = \sum \bar{X}_i^L (\beta_i^H - \beta_i^L)$$

$$U = \beta_0^H - \beta_0^L$$

$$D = C + U$$

Tendo analisado a variável dependente, segue a análise das variáveis independentes. Primeiramente o autor define o modelo estrutural, um sistema com sete equações, onde w é o salário por hora, Ed um vetor de seis *dummies* de educação, Occ um vetor de oito *dummies* de ocupação, J uma *dummy* para treinamento vocacional, M uma *dummy* para sindicatos, V uma *dummy* para veteranos do exército, T um vetor de seis *dummies* para permanência no emprego atual, B um conjunto de treze variáveis relacionadas aos antecedentes familiares e Z um conjunto de outras variáveis exógenas como idade, saúde, residência e condições do mercado de trabalho local.

$$\begin{aligned}
\log w &= f(Ed, Occ, J, M, V, T, Z) + u_1 \\
Ed &= g(Occ, V, B, Z) + u_2 \\
Occ &= h(Ed, J, V, B, Z) + u_3 \\
J &= k(Occ, V, M, B, Z) + u_4 \\
M &= l(Ed, Occ, V, B, Z) + u_5 \\
V &= m(Ed, Occ, J, B, Z) + u_6 \\
T &= n(Ed, Occ, J, M, V, Z) + u_7
\end{aligned}$$

A primeira equação é não identificada, e a única variável excluída da mesma é B . No caso, os antecedentes familiares (B) não afetam diretamente os salários, porém afetam a educação, ocupação, o treinamento vocacional, a participação sindical e se a pessoa é veterano do exército. Dessa forma, se torna difícil a exclusão de variáveis o suficiente para identificar os coeficientes. As estimativas do modelo estrutural serão viesadas, ao menos que se assuma que todos os termos de erros são não correlacionados com o erro primeira equação, u_1 . Ou seja, para obter estimativas não viesadas e utilizar mínimos quadrados ordinários (MQO) para obter os melhores estimadores é preciso assumir a hipótese restritiva $E(u_1 u_2) = E(u_1 u_3) = E(u_1 u_4) = E(u_1 u_5) = E(u_1 u_6) = E(u_1 u_7) = 0$.

Outra forma de evitar o viés seria estimar outra equação para o salário, que o autor denomina forma reduzida. Enquanto a forma estrutural considera mais amplamente as condições socioeconômicas dos indivíduos, a forma reduzida é o log do salário esperando considera apenas as condições exógenas aos indivíduos (B e Z), suas “condições de nascimento” (BLINDER, 1973, p. 442, tradução nossa).

$$\log w = F(B, Z) + v_i$$

Utilizando dados do *Michigan Survey Research Center's "Panel Study of Income Dynamics"* o autor estima R , E , C , U e D para o diferencial de salários entre homens negros e brancos e entre homens e mulheres brancos. A análise é feita tanto utilizando a forma estrutural quanto a reduzida, e os resultados são comparados. A diferença das formas estruturais das equações dos salários de homens brancos e negros mostrou uma diferença bruta de 50,8% entre os ganhos dos mesmos. 30,7%, ou seja, quase 60% dessa diferença, se da devido à diferenças

nos atributos, sendo o principal componente dessa divergência a educação, que é 14,5% maior para homens brancos. Utilizando a forma reduzida, a diferença bruta é quase a mesma (50,7%), porém a porção do diferencial atribuída aos coeficientes é de 92,2%. O coeficiente do fator idade se torna a principal causa do diferencial, assim como no caso do diferencial de salário da mulheres, que será explicado seguidamente.

O mesmo procedimento é realizado para a análise da diferença salarial entre homens e mulheres brancos. As diferenças brutas, R , encontradas utilizando a forma estrutural e a reduzida novamente são similares, sendo elas 45,6% e 45,4%, respectivamente. 113,8% do diferencial da equação estrutural e 131,7% da equação da forma reduzida é explicado pela soma de E e C , enquanto a diferença dos interceptos aponta uma vantagem para as mulheres de 68% e 86,2%, respectivamente. Desses diferenciais, E , ou seja, a parte atribuída às dotações, é relativamente irrelevante, já que 98,1% e 132% são explicados pela diferença nos atribuídas aos coeficientes na forma estrutural e reduzida, sendo 87,3% e 131,5% dela é devido ao coeficiente do fator idade em cada uma das equações.

Isto posto, é perceptível que a maior parte da diferença salarial entre sexos se da pois os salários das mulheres não aumentam durante seu ciclo de vida como acontece com o dos homens. Isso mostra que mulheres em um mesmo nível educacional e ocupacional que homens são impossibilitadas de avançar economicamente. Apesar de parte dessa diferença mostrar a dificuldade das mulheres avançarem para cargos melhores dentro de um mesmo grupo ocupacional, o autor aponta para a falha na análise de incluir explicações para a diferença na presença de homens e mulheres em diferentes ocupações (BLINDER, 1973).

O modelo de Oaxaca (1973) parte da definição de Becker (1971) de discriminação. Sendo assim, discriminação contra mulheres existe quando a diferença salarial entre homens e mulheres (W_m/W_f) é menor do que a diferença salarial que existiria em um cenário sem discriminação ($(W_m/W_f)^0$).

$$D = \frac{(W_m/W_f) - (W_m/W_f)^0}{W_m/W_f}$$

A equação acima pode ser reescrita como $\ln(D + 1) = \ln(W_m/W_f) - \ln(W_m/W_f)^0$. Como $(W_m/W_f)^0$ é desconhecido, é necessário estima-lo. O autor então utiliza uma função do logaritmo natural do salário para estimar o mesmo de homens e mulheres em função de um vetor de características individuais (Z'_i), um vetor de coeficientes (β) e do erro (u_i).

$$\ln(W_i) = Z'_i\beta + u_i$$

Dessa forma, em um caso sem discriminação, a estrutura de salários enfrentada por mulheres seria também enfrentada por homens ou a estrutura de salários enfrentada por homens também seria enfrentada por mulheres. A diferença salarial pode ser dividida em duas partes: diferenças nas características individuais discriminação. Definem G como o diferencial salarial das médias dos salários entre homens e mulheres.

$$G = \frac{\bar{W}_m - \bar{W}_f}{\bar{W}_f}$$

A equação acima pode ser reescrita como $\ln(G + 1) = \ln(\bar{W}_m) - \ln(\bar{W}_f)$, e substituindo $\ln(\bar{W}_i)$ por sua fórmula se obtém $\ln(G + 1) = \bar{Z}'_m\hat{\beta}_m + \bar{Z}'_f\hat{\beta}_f$. Como $\Delta\bar{Z}' = \bar{Z}'_m - \bar{Z}'_f$ e $\Delta\hat{\beta} = \hat{\beta}_f - \hat{\beta}_m$, é possível reescrever a equação.

$$\ln(G + 1) = \Delta\bar{Z}'\hat{\beta}_f - \bar{Z}'_m\Delta\hat{\beta}$$

Assumindo que em no caso sem discriminação a estrutura de salários das mulheres também se aplicaria aos homens, tem-se o $(W_m/W_f)^0$ abaixo.

$$\ln\left(\frac{\widehat{W}_m}{\widehat{W}_f}\right)^0 = \Delta\bar{Z}'\hat{\beta}_f$$

Com o caso sem discriminação estimado, é possível estimar os efeitos da discriminação.

$$\ln(\widehat{D} + 1) = -\bar{Z}'_m\Delta\hat{\beta}$$

Nota-se assim que o diferencial salarial ($\ln(G + 1)$) é dividido entre as diferença salarial sem discriminação, ou seja, devido a diferenças nas características individuais, e o efeito discriminação. O mesmo poderia ser feito utilizando a hipótese que a estrutura de salários dos homens se aplica para mulheres. Os resultados variam de acordo com a hipótese adotada, tornando necessário testar as regressões para os dois casos.

Partindo para o teste empírico, o autor utiliza como variáveis de controle educação (em anos), classe de trabalhador (uma *dummy* para trabalhadores sindicais, governamentais, privados ou auto-empregados), indústria (*dummy* para os diferentes tipos de indústria), ocupação (*dummy*), problemas de saúde (*dummy* igual a 1 caso o indivíduo apresente problemas de saúde e 0 caso contrário), emprego de meio período (*dummy* igual a 1 caso o indivíduo trabalhe mais de 35 horas por semana e 0 caso contrário), migração (*dummy*), estado civil (*dummy*), tamanho da área urbana, região e anos de experiência no trabalho. Como não há dados disponíveis para a última variável, é utilizada uma *proxy* para experiência potencial definida como X_i , onde A_i é a idade e E_i os anos de escolaridade do indivíduo i .

$$X_i = A_i - E_i - 6$$

A função acima é uma boa *proxy* para a experiência de homens, porém não para a de mulheres, já que elas tendem a passar mais tempo fora do mercado de trabalho, conforme será discutido no modelo de Mincer e Polachek (1974). Por isso, Oaxaca (1973) controla a também pelo número de filhos de cada mulher, que incluiria a depreciação do capital humano devido ao tempo fora do mercado. Apesar de não considerar como uma variável, o autor também controla pela raça ao fazer regressões separadas para indivíduos negros e brancos. Ao controlar por diferentes ocupações e indústrias, Oaxaca (1973) afirma eliminar alguns efeitos de discriminação presentes em barreiras de entrada para certas ocupações e indústrias, por isso também estima um modelo onde não controla por ocupação, indústria e classe de trabalhador.

O autor então faz as regressões do modelo para os dados da *Survey of Economic Opportunity* de 1967 dos Estados Unidos. Os resultados das estimações apontam para 58,4% da diferença salarial entre homens e mulheres brancos e 55,6% da diferença salarial entre homens e mulheres negros explicada por

discriminação. No entanto, é preciso analisar esses resultados cuidadosamente. Conforme o autor destaca, esse diferencial não necessariamente significa que homens e mulheres que realizam o mesmo tipo de trabalho são pagos de forma diferenciada, e sim que há uma maior concentração de mulheres em trabalhos que pagam menos (OAXACA, 1973), conforme será analisado no modelo de Fields e Wolff (1995).

Além disso, como o próprio Oaxaca (1973) destaca, o modelo residual tem o problema de não captar os *feedbacks* da discriminação no mercado sobre as características individuais, já que mulheres podem se adaptar de acordo com os vieses do mercado. No entanto nesse modelo todas as diferenças em características individuais reduzem a discriminação estimada, já que o diferencial de salários é dividido entre diferenças nas características e discriminação. Seria necessário então medir, assim como é feito com os salários, o quanto das diferenças nas características individuais existiriam em um contexto sem discriminação.

Corcoran e Duncan (1979) utilizam a decomposição de Oaxaca (1973), adicionando no modelo as jornadas dos indivíduos no mercado de trabalho (*work histories*) estudados por Mincer e Polachek (1974) e ainda controlam por mais três variáveis: treinamento no trabalho, abstenção do trabalho e imposições auto impostas sobre horas e locais de trabalho. Os autores não controlam por estado civil, apesar do mesmo ter se mostrado relevante nos demais estudos, pois procuram com as variáveis adicionadas medir o efeito direto em que o estado civil afeta o trabalho de homens e mulheres. Assim como nos demais estudos, foram feitas regressões separadas por raça. A adição dessas variáveis foi possível graças aos dados do *Panel Study os Income Dynamics*, que aplicou questionários para quase 6 mil famílias, fazendo perguntas diretas sobre essas variáveis. A crítica de Duncan (1974) sobre não levar em consideração a qualidade do treinamento, entretanto, novamente se aplica.

Os resultados da pesquisa seguiram o esperado. Homens possuem uma média de anos de escolaridade maior, sendo a média dos brancos maior que dos negros. Mulheres se mostraram muito mais suscetíveis a impor restrições sobre locais e horários de trabalho que homens. Homens completaram o dobro de treinamento no trabalho que os demais grupos.

Quando é rodada a regressão, no entanto, os coeficientes das variáveis se mostrou bastante uniforme entre os sexos e raças, de modo que homens e mulheres

recebem *payoffs* quase idênticos para as mesmas habilidades. No entanto, o termo constante da regressão de homens brancos foi muito maior que os demais, o que significa que os mesmos recebem maiores ganhos que os outros grupos em qualquer nível de habilidade. Os *payoffs* dos diferentes segmentos também se mostraram quase iguais, apenas variando significativamente entre segmentos. No entanto, o tempo que cada sexo passa em cada segmento varia em grau elevado: os homens em média passam mais tempo que as mulheres nos segmentos com maiores *payoffs*. Os autores encontram um coeficiente de depreciação de capital humano ainda menor que os encontrados por Mincer e Polachek (1974), além desses coeficientes não se mostrarem significativos.

Os resultados que a depreciação durante o tempo fora do mercado de trabalho não são significativos são contrários aos encontrados por Mincer e Polachek (1974). Abstenções e restrições auto impostas na localização e horários de trabalho também não tiveram efeitos significativos sobre o salário. Educação e histórias de trabalho (*work histories*, tradução nossa), por outro lado, mostraram efeitos significativos. Utilizando o método de Oaxaca (1973), os autores concluem que grande parte da diferença salarial não é explicada apenas pelas variáveis e fatores considerados, de modo que o mercado de trabalho não trataria todos igualmente, havendo algum tipo de discriminação.

3.2.2 Utilizando Modelo de Heckman em dois estágios

A utilização de amostras escolhidas de forma não aleatórias resulta em um viés de seleção em funções de comportamento, como as salariais ou de rendimento. Ele pode surgir devido à auto seleção por parte dos indivíduos ou dos dados, ou devido às decisões de seleção da amostra feita pelos analistas (HECKMAN, 1979). Heckman (1979) propõe uma análise desse erro como um erro de especificação, ou seja, de variáveis omitidas. Um dos problemas desse tipo de viés é que variáveis que não pertencem à função estrutural podem aparecer como estatisticamente significativas. Além disso, as fórmulas usuais para calcular o erro padrão não são mais apropriadas, e se utilizadas subestimarão os mesmo e sobrestimarão os níveis de significância (HECKMAN, 1979).

Heckman (1979) desenvolve um método para que simples funções comportamentais possam ser estimadas utilizando mínimos quadrados de forma a

evitar esse viés. Considerando um sistema com duas equações, onde X_{ji} é um vetor de regressores exógenos, β_j é um vetor de parâmetros, U_{ji} é o erro.

$$\begin{aligned} Y_{1i} &= X_{1i}\beta_1 + U_{1i} \\ Y_{2i} &= X_{2i}\beta_2 + U_{2i} \end{aligned}$$

Supõe-se então que se deseja estimar $Y_{1i} = X_{1i}\beta_1 + U_{1i}$, porém há dados faltando para Y_{1i} , e que há dados disponíveis para Y_{1i} se $Y_{2i} \geq 0$, e caso $Y_{2i} < 0$ não há dados para Y_{1i} . Com essas suposições, é escrita a função de regressão da subamostra.

$$E(Y_{1i}|X_{1i}, Y_{2i} \geq 0) = X_{1i}\beta_1 + E(U_{1i}|U_{2i} \geq -X_{2i}\beta_2)$$

Assumindo que a densidade $h(U_{1i}, U_{2i})$ é bivariada e que σ_{ij} representa o desvio padrão.

$$\begin{aligned} E(U_{1i}|U_{2i} \geq -X_{2i}\beta_2) &= \frac{\sigma_{12}}{(\sigma_{22})^{\frac{1}{2}}} \lambda_i \\ E(U_{2i}|U_{1i} \geq -X_{2i}\beta_2) &= \frac{\sigma_{22}}{(\sigma_{22})^{\frac{1}{2}}} \lambda_i \end{aligned}$$

Sendo ϕ e Φ a densidade e a distribuição da função para uma variável normal e λ_i a função monótona decrescente da probabilidade de uma observação ser selecionada para a amostra.

$$\begin{aligned} \lambda_i &= \frac{\phi(Z_i)}{1 - \Phi(Z_i)} = \frac{\phi(Z_i)}{\Phi(-Z_i)} \\ Z_i &= -\frac{X_{2i}\beta_2}{(\sigma_{22})^{\frac{1}{2}}} \end{aligned}$$

Desse modo, o modelo estatístico completo para perturbações populacionais normais pode ser desenvolvido.

$$\begin{aligned} Y_{1i} &= E(Y_{1i}|X_{1i}, Y_{2i} \geq 0) + V_{1i} \\ Y_{2i} &= E(Y_{2i}|X_{2i}, Y_{2i} \geq 0) + V_{2i} \end{aligned}$$

Se Z_i e conseqüentemente λ_i são conhecidos, é possível acrescentar λ_i como um regressor na equação $Y_{1i} = E(Y_{1i}|X_{1i}, Y_{2i} \geq 0) + V_{1i}$ e estimar a mesma utilizando MQO. Se os mesmos não são conhecidos e não se tem dados para Y_{1i} se $Y_{2i} \leq 0$, porém se conhece X_{2i} quando $Y_{2i} \leq 0$, é possível estimar λ_i . Primeiramente é necessário estimar os parâmetros da probabilidade de $Y_{2i} \leq 0$, ou seja, $\frac{\beta_2}{(\sigma_{22})^{1/2}}$, utilizando uma análise *probit* para toda a amostra. Tendo estimado $\frac{\beta_2}{(\sigma_{22})^{1/2}}$ é possível estimar Z_i e conseqüentemente λ_i , e todos esses estimadores são consistentes. O λ_i estimado, $\hat{\lambda}_i$, pode então ser acrescentado como um regressor na equação $Y_{1i} = E(Y_{1i}|X_{1i}, Y_{2i} \geq 0) + V_{1i}$ e desse modo eliminar o viés de seleção (HECKMAN, 1979).

Aplicando o método acima para a equação salarial, temos $W_i = x'_i\beta + \varepsilon_i$, onde W_i é o salário, x'_i as variáveis observáveis de cada indivíduo i e ε_i o termo de erro (HUDSON, 2013² *apud* PEREIRA; OLIVEIRA, 2013, p. 1). Seguindo o exposto acima, tem-se que W_{2i} é observável apenas para $W_{1i} > 0$ (CAMERON; TRIVEDI, 2005³ *apud* PEREIRA; OLIVEIRA, 2013, p. 2)

$$W_{2i} = \{1 \text{ se } W_{1i} > 0, 0 \text{ se } W_{1i} \leq 0\}$$

Então é adicionada a estimativa para $\lambda_i(z_i\gamma)$, o regressor omitido, para valores positivos de W_{2i} , onde $\hat{\beta}_1$ é obtido através da regressão *probit* e $\hat{\lambda} = \lambda(x_{1i}\hat{\beta}_1)$. Assim o estimador de β_2 é consistente e, apesar de uma perda de eficiência em comparação com o estimador de máxima verossimilhança sob normalidade conjunta dos erros, ele requer suposições distribucionais mais fraca que o mesmo (CAMERON; TRIVEDI, 2005⁴ *apud* PEREIRA; OLIVEIRA, 2013, p. 3).

$$W_{2i} = x'_{2i}\beta_2 + \sigma_{12}\lambda(x'_{1i}\hat{\beta}_1) + \varepsilon_i$$

² HUDSON, John. **Econometrics Lecture: Heckman's Sample Selection Model**. Bath University – UK. Disponível em: <staff.bath.ac.uk/hssjrh/Heckman_LECTURE.doc>. Acesso em: 03 janeiro 2013.

³ CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: Methods and Applications**. New York: Cambridge University Press, 2005.

⁴ CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: Methods and Applications**. New York: Cambridge University Press, 2005.

Assim, se propõe duas equações com as características dos indivíduos: uma determina se o mesmo participa do mercado e a outra seu rendimento. A primeira equação é a equação de seleção, que determina a decisão do indivíduos de entrar ou não no mercado de trabalho (PEREIRA; OLIVEIRA, 2013).

$$L_i = \alpha_1 + \alpha_2 \text{sexo}_i + \alpha_3 \text{raça}_i + \alpha_4 \text{anosdeestudo}_i + \alpha_5 \text{experiência}_i + \alpha_6 \text{experiência}_i^2 + \alpha_7 \text{zonaurbana}_i + \varepsilon_i$$

Onde L_i é uma *dummy* com valor 1 se o indivíduo participa do mercado de trabalho e 0 caso contrário. A segunda equação busca estimar os rendimentos, onde representa o logaritmo do rendimento por hora.

$$\ln W = \alpha_1 + \alpha_2 \text{experiência}_i + \alpha_3 \text{experiência}_i^2 + \alpha_4 \text{raça}_i + \alpha_5 \text{niveldeescolaridade}_i + \alpha_7 \text{exoilarid}$$

Utilizando esse modelo, os autores afirmam que conseguem resultados mais robustos dos que os obtidos por MQO, e o regressor extra lambda se mostrou significativo e com o sinal negativo. Os autores utilizam os dados do Censo de 2010 para o Rio Grande do Sul para rodar a regressão e após realizar a decomposição de Oaxaca, calculando a as previsões no caso da estrutura de rendimentos das mulheres se aplicar para ambos os sexos e para a estrutura de rendimentos dos homens se aplicar a ambos os sexos. Seus resultados mostram uma diferença de R\$ 1,15 real por hora de diferença entre o rendimento dos homens e das mulheres. O coeficiente positivo do componente residual da diferença salarial indica que há discriminação presente no mercado (PEREIRA; OLIVEIRA, 2013).

É importante atentar para alguns pontos do estudo de Pereira e Oliveira (2013). Primeiramente, apesar da maior robustez dos estimadores afirmada pelos mesmo e da significância do regressor λ , os valores encontrados são muito similares quando comparados com os encontrados utilizando MQO. Além disso, os autores utilizam a mesma estimação de experiência para homens e mulheres, que como foi visto anteriormente e será mais discutido na sequência, pode trazer problemas pois as mulheres tendem a ter mais períodos de abstenção do mercado que os homens. Por último, diferentemente dos demais autores analisados, Pereira e Oliveira (2013)

não consideram a educação como uma função quadrática em nenhuma das duas equações, fazendo isso apenas para a experiência.

Ademais, conforme sugerido acima, a decomposição de Oaxaca deixa os resultados abertos para interpretação, já que o resíduo explicado pela discriminação pode indicar discriminação do empregador, pré condicionamento cultural ou diferenças não observadas em produtividade (COX; NYE, 1989). Outro problema com o modelo descrito acima são os micro dados disponíveis. Mesmo em uma situação ideal onde se possa controlar por um número grande de fatores, é muito improvável que hajam dados sobre a produtividade do indivíduo como força física ou energia (COX; NYE, 1989).

3.3 ESTIMAÇÃO DE PRODUTIVIDADE DE COX E NYE

O modelo de Cox e Nye (1989) utiliza dados da indústria têxtil francesa do século XIX. O grande diferencial desse modelo em relação aos demais é que ele busca estimar o diferencial em produtividade, utilizando uma função de produção. Os autores partem de uma definição usual de lucro (π), onde P é o preço, Q a quantidade, L_m e L_f o número de homens e o número de mulheres empregados, respectivamente, K o capital, r a remuneração do capital e w_m e w_f os salários dos homens e das mulheres, respectivamente.

$$\pi = PQ(L_m, L_f, K) - w_m L_m - w_f L_f - rK$$

Sendo $\frac{d\pi}{dL_f}$ a discriminação do empregador, a condição de primeira ordem para a maximização do salário é descrita abaixo. Caso não houver discriminação, o salário das mulheres será igual ao seu produto marginal (ou seja, igual ao preço em um mercado competitivo).

$$PQ_f - \frac{d\pi}{dL_f} = w_f$$

Assumem então que a função de produção tem o formato de uma Cobb-Douglas, com β_1 , β_2 e β_3 sendo as elasticidades e A o termo constante.

$$\ln Q = \ln A + \beta_1 \ln L_m + \beta_2 \ln L_f + \beta_3 K$$

Se os empregadores se importarem apenas com o lucro (isto é, $\frac{d\pi}{dL_f} = 0$), então a suposição de Cobb-Douglas em que a relação entre as elasticidades dos rendimentos se iguala a relação dos custos de trabalho se mantém.

$$\frac{\beta_1}{\beta_2} = \frac{w_m L_m}{w_f L_f}$$

Utilizando dados sobre as matérias primas, empregados, horas trabalhadas e potência das máquinas, Cox e Nye (1989) estimam a função de produção descrita acima para algodão e lã. Para testar a hipótese nula de não discriminação, é adicionada a condição de não discriminação ($\frac{\beta_1}{\beta_2} = \frac{w_m L_m}{w_f L_f}$) e aplicado um teste F. Os resultados diferiram para os dois períodos analisados. Para os dados do período entre 1860 e 1865 não foi possível rejeitar a hipótese de não-discriminação nem a hipótese de igualdade dos produtos marginais entre os sexos. Os testes feitos para o período 1839-45 novamente não rejeitam a hipótese nula de não-discriminação pelos empregadores, porém para esse período as elasticidades do produto são uniformemente menores para mulheres do que para homens e o padrão da relação entre a produtividade entre homens e mulheres e sua relação de salários segue a prevista pela teoria econômica, isto é, uma discrepância maior entre produtividades de homens e mulheres estão associados com maiores diferenças salariais.

Além de ser o único dos modelos analisados que estima diferenças na produtividade, o modelo de Cox e Nye (1989) também é o único que não encontra evidências de discriminação salarial. No entanto, como destacam os autores, sua análise não incorpora segregação ocupacional, tanto no modelo quanto nos dados, que são coletados apenas para um tipo de indústria.

3.4 MODELO DE DISCRIMINAÇÃO INTERINDUSTRIAL

Um modelo cujo foco é justamente determinar quais indústrias possuem as maiores e menores diferenças salariais é o modelo de discriminação interindustrial desenvolvido por Fields e Wolff (1995). Eles apontam que o principal motivo para os

maiores valores encontrados para os diferenciais salariais e entre indústrias é o menor nível de agregação de indústrias utilizados pelos mesmos.

Primeiramente, é definida a função salarial, onde W_i é o salário por hora $EDUC_i$ representa a quantidade de anos de estudo, EXP_i representa a experiência no mercado de trabalho, e é estimada pela idade menos os anos de estudo menos 6, $URBAN_i$ é uma *dummy* para residências urbanas (em cidades centrais), $SMSA_i$, do inglês *Standard Statistical Metropolitan Area*, representa o tamanho da população da região, $REGION_{ij}$ é uma *dummy* para as diferentes regiões do país, $MARRIED_i$ é uma *dummy* igual a 1 se o indivíduo é casado, $RACE_i$ uma *dummy* para raça (não-branco e branco) e $OCCUP_{ij}$ um conjunto de 13 *dummies* de ocupação. A variável de maior interesse é a $INDUS_{ji}$. Os autores utilizam três conjuntos diferentes de *dummies* para a mesma: um com indústrias de um dígito (14 indústrias), outro com indústrias de dois dígitos (46 indústrias) e por fim um com indústrias de três dígitos (224 indústrias).

$$\ln W_i = a + b_1 EDUC_i + b_2 EXP_i + b_3 EXP_i^2 + b_4 URBAN_i + b_5 SMSA_i + \sum \zeta_j REGION_{ij} \\ + b_7 MARRIED_i + b_8 RACE_i + \sum \alpha_j OCCUP_{ij} + \sum \beta_i INDUS_{ji} + \varepsilon_i$$

A amostra é dividida em duas, uma contendo apenas trabalhadores com trabalho fixo durante o ano e outra com todos os trabalhadores. São feitas regressões separadas por sexo utilizando MQO. Os autores seguem a definição de diferencial salarial entre indústrias, d_j , de Krueger e Summers (1988 *apud* FIELDS; WOLFF, 1995), onde $\hat{\beta}_j$ é o coeficiente estimado para a *dummy* da indústria j e s_k é a proporção da amostra empregada na indústria k . Os diferenciais foram normalizados como desvios de uma média ponderada pelo emprego dos coeficientes da regressão.

$$d_j = \hat{\beta}_j - \left[\sum_k \hat{\beta}_k s_k \right]$$

Os resultados mostram uma dispersão em d_j tão grande para homens quanto para mulheres. Além disso, utilizando níveis menos desagregados de indústrias resultam em dispersões maiores, assim como a divisão das regressões por gênero

também revelam uma maior dispersão de d_j . Por exemplo, os autores utilizam o CPS (*Current Population Survey*) de março de 1988 dos Estados Unidos e encontram uma dispersão de 0,45 em d_j para a amostra de mulheres com trabalho fixo durante todo ano para indústrias de um dígito, já para essa mesma amostra para indústrias de três dígitos, ou seja, uma desagregação maior, a dispersão encontrada é de 1,38 (FIELDS; WOLFF, 1995).

Mesmo controlando pelas diversas variáveis citadas anteriormente, os coeficientes da maioria das *dummies* das indústrias se mostraram significativos. Além disso, a presença das mesmas aumenta os coeficientes de determinação das equações. As *dummies* de indústria de um dígito explicam sozinhas 12% da variação de rendimentos entre trabalhadores, enquanto as variáveis de indústria de três dígitos explicam 22% dessa variação (FIELDS; WOLFF, 1995).

As variações salariais interindustrial entre gêneros são altamente correlacionadas, apesar dessa correlação diminuir conforme o nível de agregação de indústrias diminui (de 0,95 para indústrias de um dígito para 0,79 para indústrias de três dígitos). Apesar das semelhanças nos resultados das regressões para as equações salariais interindustriais entre para homens e mulheres (uma dispersão similar, assim como um ranking parecido entre as indústrias com maior e menor diferencial salarial por indivíduo separada por gênero), os coeficientes estimados possuem diferenças importantes (FIELDS; WOLFF, 1995).

Para mostrar as mesmas, os autores utilizam o teste t e o teste de Chow. Utilizando o teste t, cerca de 40% das diferenças entre os coeficientes dos homens e mulheres são estatisticamente significativas a um nível de 10%. O teste de Chow primeiramente une homens e mulheres em uma única amostra. Então compara a soma dos erros quadrados entre as equações com um único conjunto de *dummies* para indústrias (mas as demais variáveis separadas para homens e mulheres) e outra equação como todas as variáveis separadas por gênero. Os resultados apontam para diferenças significativas nos coeficientes de homens e mulheres (FIELDS; WOLFF, 1995).

Os autores então definem o diferencial entre salários por indústria e o diferencial salarial geral entre gêneros. O primeiro é definido como $g_j = (\hat{\beta}_j^f - \hat{\beta}_j^m) + (\hat{\alpha}^f - \hat{\alpha}^m)$, onde $\hat{\beta}_j^f$ é os coeficientes da equação salarial das mulheres e $\hat{\beta}_j^m$ dos homens, e $\hat{\alpha}^f$ é a constante da equação salarial das mulheres e $\hat{\alpha}^m$ dos homens. Os

diferenciais encontrados variam de acordo com o nível de agregação da indústria, com a amplitude indo de 0,17 (0,4 em agricultura, onde, de acordo com o g_j calculado as mulheres são recebem mais que homens, até -0,13 em vendas em varejo) para níveis maiores de agregação, até 0,6 para o menor índice de variação. Para o cálculo do diferencial salarial geral entre gêneros, a equação dos salários é dividida entre homens e mulheres.

$$\ln W_i^f = \alpha^f + X_i^f \theta^f + \sum_{j=2}^J \beta_j^f s_{ij}^f + \varepsilon_i^f$$

$$\ln W_i^m = \alpha^m + X_i^m \theta^m + \sum_{j=2}^J \beta_j^m s_{ij}^m + \varepsilon_i^m$$

A primeira equação é a regressão para a amostra de mulheres e a segunda para a amostra de homens. Os sobescritos f e m se referem a mulheres e homens, respectivamente. X_i é um vetor de características, s_{ij} é uma variável binária que é igual a 1 se o trabalhador i está empregado na indústria j dentro de um total de J indústrias, ε_i é o erro do tipo ruído branco e α , β_j e θ são parâmetros a serem estimados.

A diferença salarial geral é apenas a diferença entre as duas equações (HORRACE; OAXACA, 2001). A mesma pode ser decomposta em três partes: a diferença devido à diferente distribuição de homens e mulheres entre as indústrias, ou seja, $\sum_j \bar{\beta} (s_{ij}^m - s_{ij}^f)$; a diferença devido aos diferentes coeficientes estimados para as indústrias de homens e mulheres, $\sum_j \bar{s}_j (\beta_j^m - \beta_j^f)$; e por fim a parte explicada pelas demais variáveis. A diferença total encontrada por Fields e Wolff (1995) foi entre 34% e 36%, dependendo da amostra utilizada, e do total dessa diferença de 13 à 19% são explicados pela diferença na distribuição de homens e mulheres entre indústrias, com o percentual variando de acordo com o nível de agregação das indústrias. Isso significa que a distribuição de mulheres tende a ficar mais concentrada em indústrias que pagam menos (FIELDS; WOLFF, 1995).

Horrace e Oaxaca (2001) apontam dois problemas com o modelo acima. Primeiramente, os autores não divulgaram os erros padrões, que são importantes para inferências sobre a significância estatística de diferenças salariais individuais e sobre a magnitude relativa entre diferenças salariais. Além disso, a estimação da

diferença salarial $\widehat{g}_j = (\widehat{\beta}_j^f - \widehat{\beta}_j^m) + (\widehat{\alpha}^f - \widehat{\alpha}^m)$ varia de acordo com o grupo deixado de fora nas variáveis binárias contidas em x_i , como por exemplo raça, ou seja, \widehat{g}_j varia se omitimos o grupo “branco” ou o grupo “não branco”. Hoxby e Oaxaca (2001) então propõe mudanças na função de discriminação do modelo de modo a evitar esse problema de identificação. Ao invés de \widehat{g}_j , seria utilizado $\widehat{\phi}_j$.

$$\widehat{\phi}_j = (\widehat{\beta}_j^f - \widehat{\beta}_j^m) + (\widehat{\alpha}^f - \widehat{\alpha}^m) + \overline{x}_j^f (\widehat{\theta}^f - \widehat{\theta}^m)$$

Com esse novo estimador o problema de identificação seria evitado, pois mudanças em $(\widehat{\alpha}^f - \widehat{\alpha}^m)$ seriam compensadas por mudanças em $(\widehat{\theta}^f - \widehat{\theta}^m)$. Esse estimador, todavia, ainda varia de acordo com a média das características por indústria (\overline{x}_j^f). Adotando uma média para a amostra de mulheres em todas as indústrias (ou seja, da amostra total), se obtém um estimador $\widehat{\delta}_j$ que não varia por indústria.

$$\widehat{\delta}_j = (\widehat{\beta}_j^f - \widehat{\beta}_j^m) + (\widehat{\alpha}^f - \widehat{\alpha}^m) + \overline{x}^f (\widehat{\theta}^f - \widehat{\theta}^m)$$

Para evitar o problema de invariância contido no intercepto, os autores também calculam a diferença salarial relativa por indústria. Utilizando os mesmos dados que Fields e Wolff (1995), os autores repetem o modelo estimando o log do salário por hora como função de uma constante, educação, experiência potencial, experiência potencial ao quadrado, residência urbana (variável binária), número de habitantes por residência, estado civil (variável binária), raça (variável binária), ocupação (doze variáveis binárias) e indústria (treze variáveis binárias) e estimando g_j , ϕ_j , δ_j e a diferença salarial relativa por indústria.

Como esperado, os resultados variam de acordo com o estimador utilizado. Os resultados de \widehat{g}_j não são todos negativos, assim como quando calculados por Fields e Wolff (1995), enquanto os de $\widehat{\phi}_j$ são, o que significa que mulheres recebem menos em todas as indústrias, além de serem todos estatisticamente significativos a um nível de 95%. O ranque também varia dependendo do estimador utilizado. Esses resultados provam que diferenças em características pessoais por indústria podem mascarar os efeitos da própria indústria na diferença salarial entre homens e mulheres. Ainda assim, as magnitudes de diferenças salariais entre indústrias é pequena e os ranques de indústrias não são estatisticamente confiáveis (HORRACE; OAXACA, 2001).

4 ESTIMAÇÃO EMPÍRICA: DIFERENÇAS EM RENDIMENTOS POR GÊNERO NA REGIÃO METROPOLITANA DE PORTO ALEGRE

Nos capítulos anteriores foram apresentados os modelos e teorias econômicas sobre discriminação, que são utilizados como base teórica do presente capítulo. Este busca fazer um estudo empírico da diferença salarial da região metropolitana de Porto Alegre. Para tal, primeiramente é descrita a base de dados utilizada e a amostra retirada da mesma, para depois passar para a modelagem e resultados feita com o software R. O modelo escolhido, dos apresentados no capítulo anterior, foi o de Oaxaca-Blinder, com alterações nas variáveis utilizadas, que também é explicada neste capítulo.

4.1 A BASE DE DADOS

A base de dados utilizada é a Pesquisa de Emprego e Desemprego da Região Metropolitana de Porto Alegre (PED) do ano de 2017, feita pela FEE. A coleta dos dados é feita por meio de questionários aplicados à unidades domiciliares, totalizando 8.100 domicílios. A seleção dos domicílios é feita em dois estágios: primeiramente são sorteados os setores censitários (501 ao total) e são listados todos os domicílios dos setores selecionados. Em seguida, são selecionados os domicílios a serem entrevistados.

A amostra inicial disponibilizada pela FEE possui 38.104 observações, cada uma correspondente a um indivíduo, e 118 variáveis. Desta amostra inicial, são removidos indivíduos que não pertencem ao mercado de trabalho e aqueles cujos dados não podem ser utilizados na análise, conforme será explicado. A subamostra resultante, que será utilizada para a modelagem e análise das diferenças salariais, é de 7.107 observações.

Essa subamostra foi obtida a partir de alterações na amostra do PED. Primeiramente foram retiradas as pessoas com menos de 10 anos, inativos puros e cuja situação é indefinida (as com valores 7, 6 e 0, respectivamente, da variável SIT – Situação Ocupacional), já que as mesmas não se encontram no mercado de trabalho.

Além disso, foram removidos os indivíduos sem observações na variável renda do trabalho principal (Q422 – Quanto o Sr.(a) Ganhou pelo Trabalho Realizado no Mês Passado – Remuneração Líquida do Trabalho Principal), assim como aqueles cuja renda principal é zero, “não se aplica” e “sem declaração” (valor da variável igual à 10.000.000 e 10.000.001, respectivamente). Também foram removidos os indivíduos com zero horas trabalhadas efetivamente no trabalho principal na semana anterior (Q431 – Quantas Horas o Sr.(a) Trabalhou Efetivamente em cada um dos seus Trabalhos na Semana Passada? – No Trabalho Principal). Também foi removido o único indivíduo cujo valor da variável C102 – Última Série Concluída (Grau) era igual à 10, significando não se aplica. Nenhum indivíduo da amostra restante possuía essa variável igual a 10, que significa sem declaração.

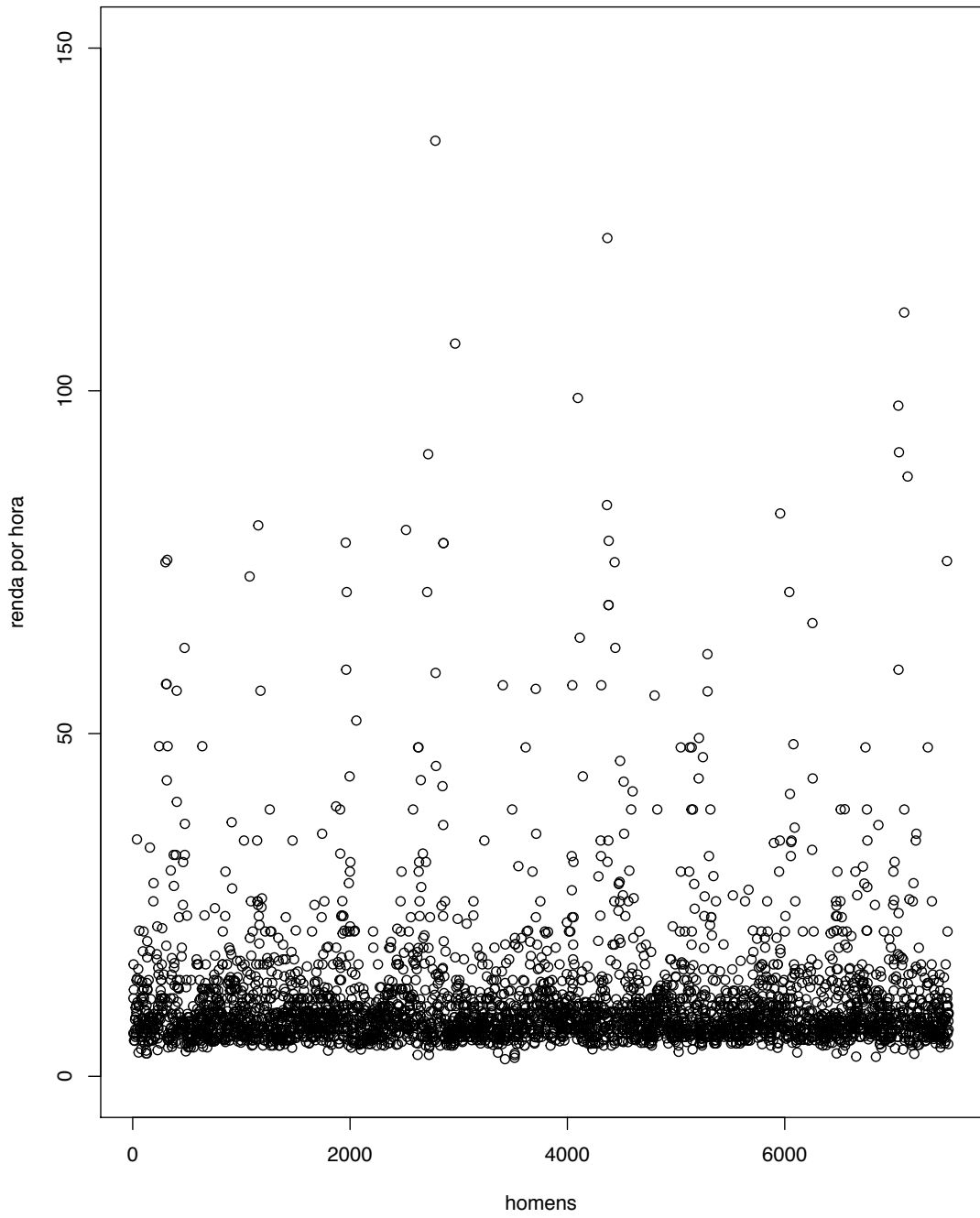
Por fim, foram removidos os 22 indivíduos da amostra obtida que pertencem ao grupo dos trabalhadores agropecuários, florestais, da caça e pesca e os 2 indivíduos que não sabem a que setor pertence a empresa em que trabalham, para evitar problemas de colinearidade, conforme explicitado no capítulo 4.3.

4.2 ANÁLISE DESCRITIVA DOS DADOS

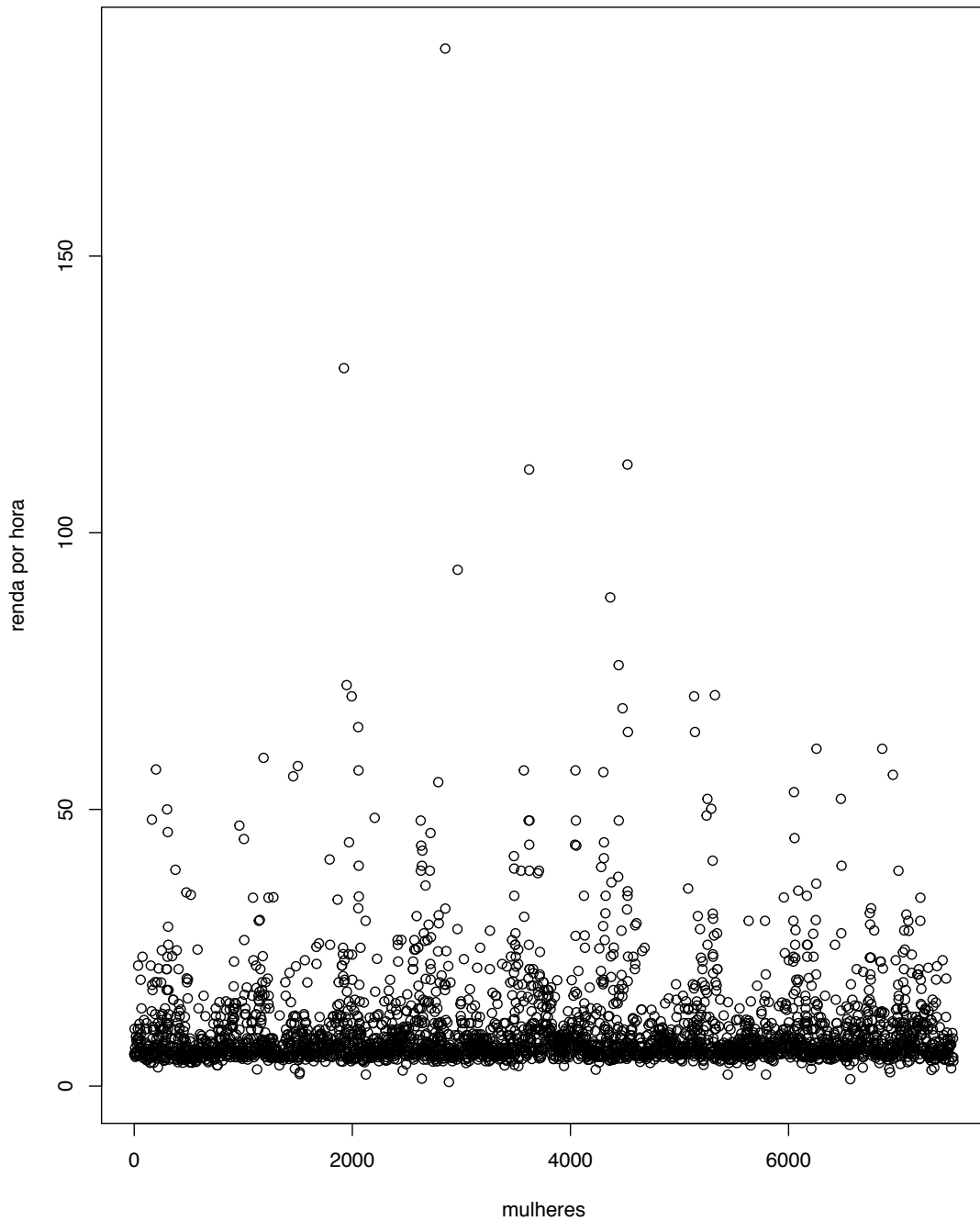
A amostra analisada possui 7.107, sendo 3.798 homens e 3.309 mulheres. Destes, 2.683, ou seja 37,75%, são casados e 6.093 (85,73%) são brancos. A renda média por hora é R\$ 10,56, e a média de horas trabalhadas semanalmente é 41,17 no trabalho principal e 0,28 no trabalho adicional. As mulheres possuem uma média de horas trabalhadas no trabalho principal menor que os homens, sendo elas 40,07 e 42,14, respectivamente. O contrário acontece com a média de horas do trabalho adicional: as mulheres trabalham em média 0,34 horas no mesmo, enquanto os homens trabalham 0,22.

Ao analisar a dispersão da renda por hora dos dois sexos, é perceptível que mais homens possuem rendas maiores. A concentração dos mesmos na faixa de renda entre 50 e 100 reais por hora é maior que a das mulheres. Também nota-se que a maior concentração de indivíduos de ambos os sexos é em torno da renda média de R\$ 10,00 por hora.

Gráfico 1 - Dispersão da Renda dos Homens



Fonte: Elaboração própria (2018).

Gráfico 2. Dispersão da Renda das Mulheres

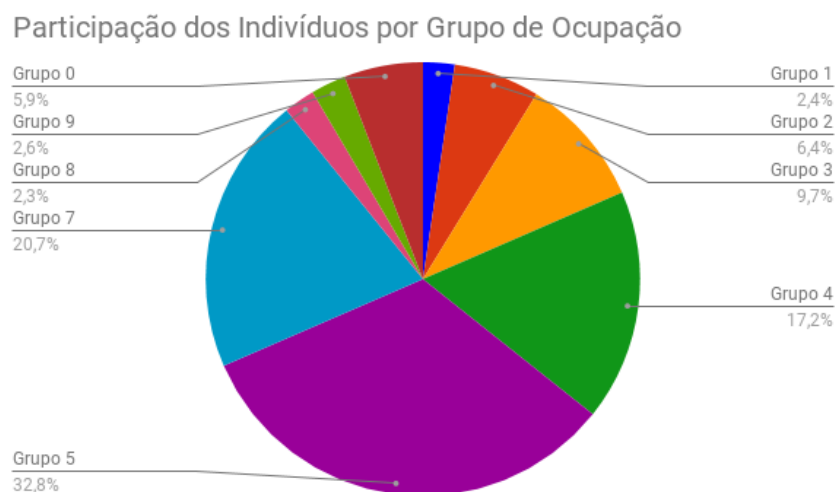
Fonte: Elaboração própria (2018).

A média de anos de educação para a amostra é de 10,85 anos, sendo que as mulheres possuem uma média maior que os homens (11,22 e 10,53 anos, respectivamente). O contrário ocorre com anos de experiência, cuja média da amostra é 5,22 anos. Para os homens, essa média é de 5,44 anos, enquanto para as mulheres é 4,97 anos.

As ocupações estão listadas na base de dados de acordo com a Classificação Brasileira de Ocupações (CBO) de 2002. Os dados foram agregados nos grandes grupos da própria CBO: grupo 0 (forças armadas, policiais e bombeiros militares), grupo 1 (membros superiores do poder público, dirigentes de organizações de interesse público e de empresas e gerentes), grupo 2 (profissionais das ciências e das artes), grupo 3 (técnicos de nível médio), grupo 4 (trabalhadores de serviços administrativos), grupo 5 (trabalhadores de serviços, vendedores do comércio em lojas e mercados), grupo 6 (trabalhadores agropecuários, florestais, da caça e pesca), grupo 7 (trabalhadores da produção de bens e serviços industriais de sistemas de produção discretos e que lidam mais com a forma do produto), grupo 8 (trabalhadores de produção de bens e serviços industriais de sistemas de produção que são contínuos, ou seja, físico-químicos) e grupo 9 (trabalhadores de manutenção e reparo) (BRASIL, 2017).

Analisando a participação dos indivíduos nos diferentes grupos de ocupação, se constata que os grupos 5 (trabalhadores dos serviços, vendedores do comércio em lojas e mercados), 7 (trabalhadores da produção de bens e serviços industriais) e 4 (trabalhadores de serviços administrativos) são os que ocupam um maior número de indivíduos, conforme ilustrado pelo gráfico 3.

Gráfico 3 - Participação dos Indivíduos por Grupo de Ocupação



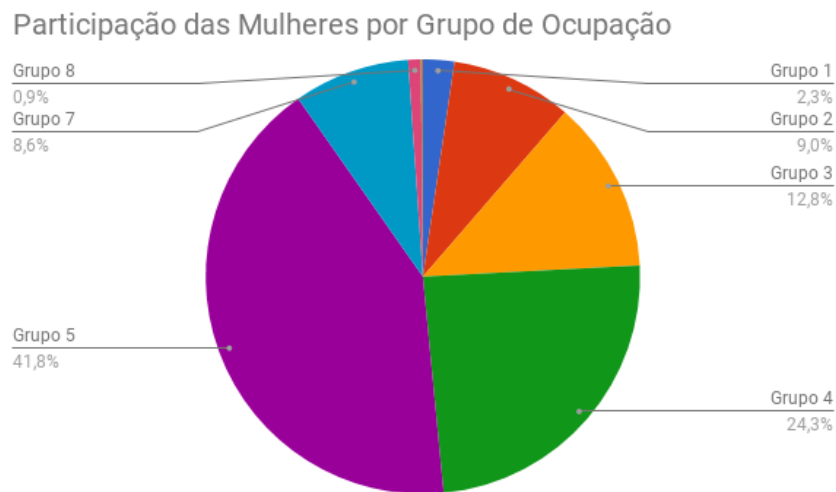
Fonte: Elaboração própria (2018).

Essa distribuição muda de acordo com o sexo. O grupo 5 ocupa quase a metade das trabalhadoras (41,8%), e o grupo 7 passa a ter uma participação menor

que o grupo 4, de acordo com o gráfico 4. Já para os homens o grupo 7 é o mais relevante, seguido pelo grupo 5 que novamente ocupa um alto percentual de trabalhadores, como ilustra o gráfico 5.

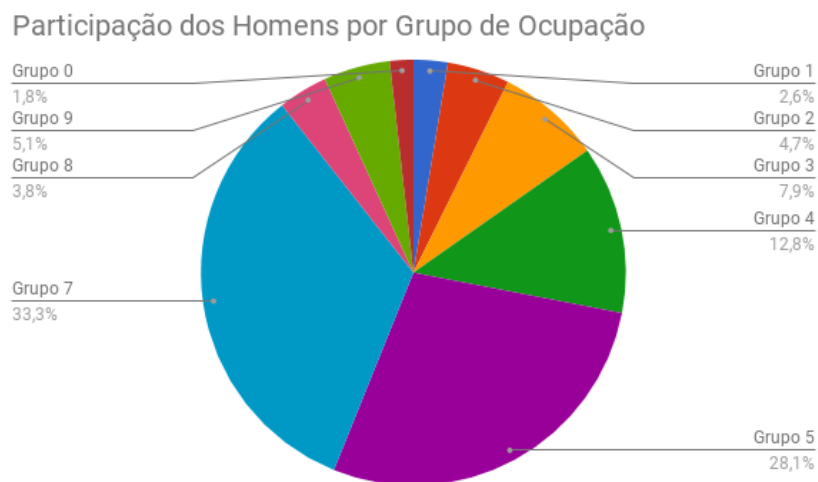
Dessa forma, se conclui que os serviços possuem grande relevância nas ocupações dos trabalhadores de ambos os sexos, mas principalmente para mulheres, na região metropolitana de Porto Alegre. A ocupação de trabalhador de bens e serviços industriais é relativamente mais importante para o emprego masculino do que o feminino, empregando uma grande quantidade de homens (um terço da amostra masculina) e pequena de mulheres (8,6% das mulheres da amostra).

Gráfico 4 - Participação das Mulheres por Grupo de Ocupação



Fonte: Elaboração própria (2018).

Gráfico 5 - Participação dos Homens por Grupo de Ocupação

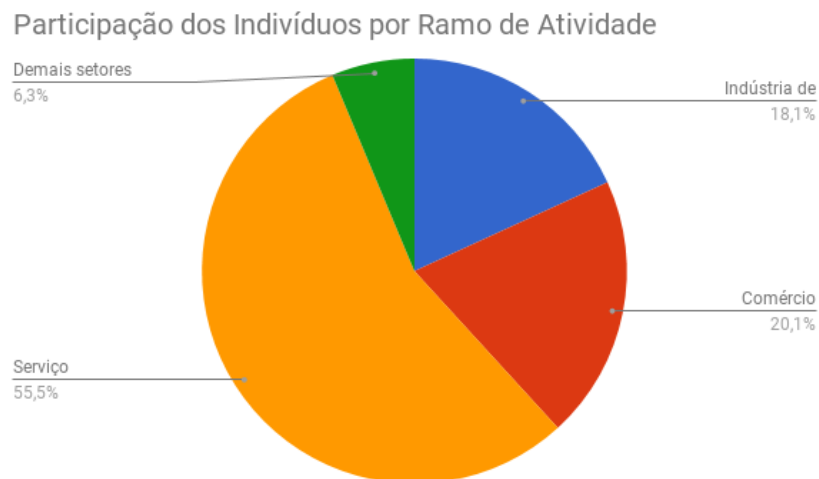


Fonte: Elaboração própria (2018).

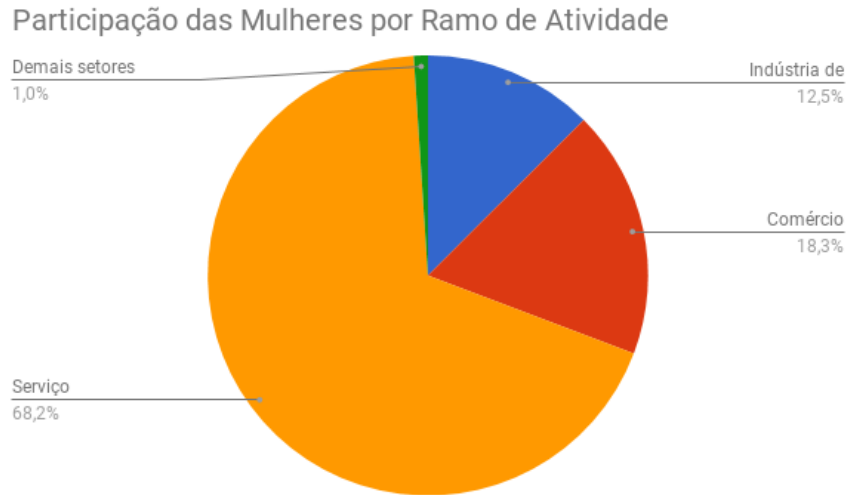
Observando o gráfico 6, que ilustra a participação dos indivíduos por ramo de atividade, se percebe que o setor emprega a maioria (55%) dos trabalhadores da amostra, em congruência com a análise anterior. Quando apenas a amostra feminina é considerada, esse percentual aumenta para 68,2%, enquanto para os homens ela diminui (para 44,3%, sendo ainda o ramo que mais emprega).

Na comparação entre as amostras masculinas e femininas, também se destaca o fato de o setor industrial empregar mais homens, conforme explicita o gráfico 8. O percentual dos homens empregados pelo ramo industrial excede a participação do ramo quando é considerada toda a amostra, mostrando a relativa importância para o emprego masculino. Já o ramo comercial emprega um número percentual similar de homens e mulheres da amostra.

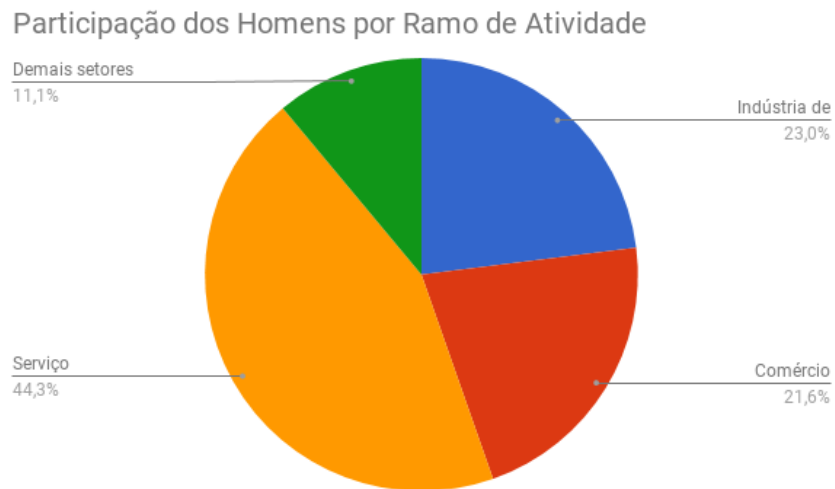
Gráfico 6 - Participação dos Indivíduos por Ramo de Atividade



Fonte: Elaboração própria (2018).

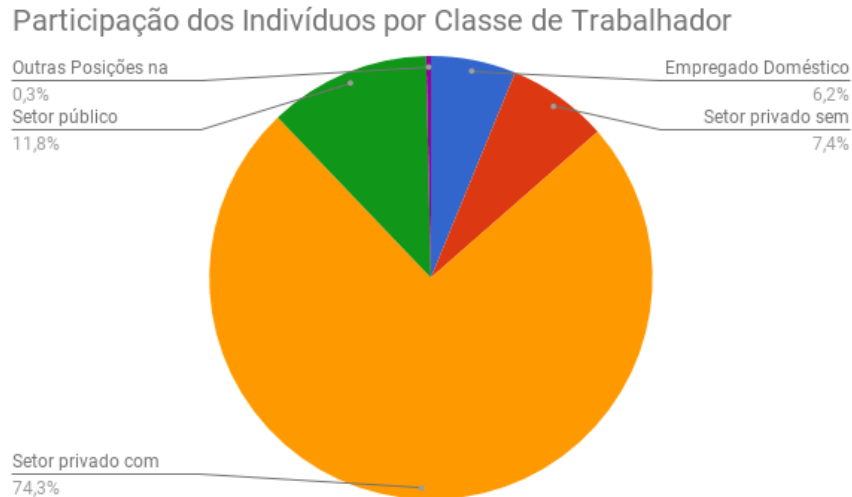
Gráfico 7 - Participação das Mulheres por Ramo de Atividade

Fonte: Elaboração própria (2018).

Gráfico 8 - Participação dos Homens por Ramo de Atividade

Fonte: Elaboração própria (2018).

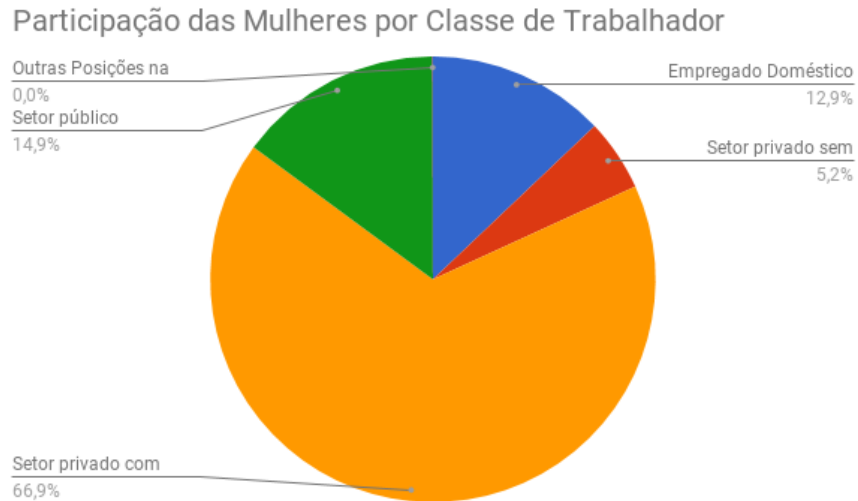
Por último é analisada a participação percentual do total da amostra e de homens e mulheres da amostra separadamente por classe de trabalhador. Se destaca o fato de trabalhadores do setor privado com carteira assinada ser a principal classe em todos os três casos, empregando 74,3%, 80,7% e 66,9% do total de trabalhadores, dos homens e das mulheres da amostra, respectivamente.

Gráfico 9 - Participação dos Indivíduos por Classe de Trabalhador

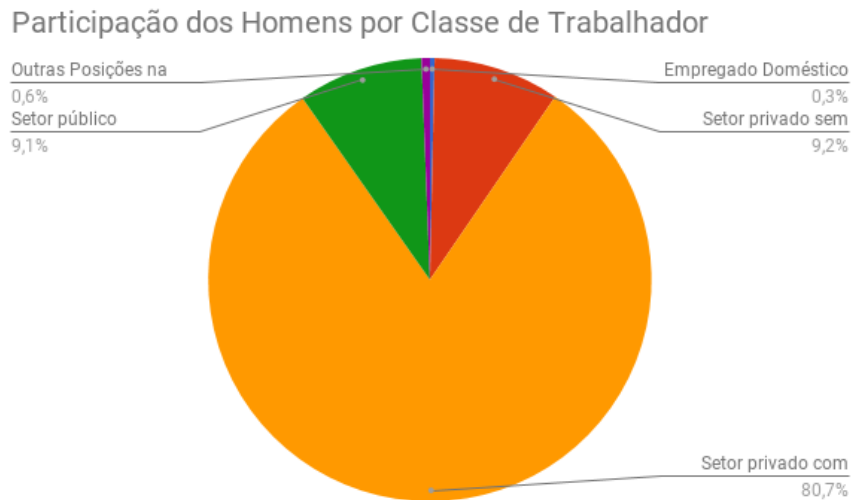
Fonte: Elaboração própria (2018).

Outro elemento que se destaca é a participação do emprego doméstico para os dois sexos. Enquanto é a classe de trabalhador de 12,9% das mulheres da amostra (428 mulheres) conforme o gráfico 8, corresponde a classe de trabalhador de apenas 0,3% dos homens (13 homens), conforme o gráfico 9. Dessa forma, o emprego doméstico é uma classe importante de trabalhador para as mulheres e apenas marginal para os homens, sendo enquadrados nela menos que homens que em “outras posições” (onde estão classificados 23 homens e apenas uma mulher).

A participação percentual de homens no setor privado sem carteira assinada é o dobro da de mulheres. Já a participação de mulheres no setor público é maior que a dos homens, sendo essas 14,9% e 9,1%, respectivamente.

Gráfico 10 - Participação das Mulheres por Classe de Trabalhador

Fonte: Elaboração própria (2018).

Gráfico 11 - Participação dos Homens por Classe de Trabalhador

Fonte: Elaboração própria (2018).

4.3 AS VARIÁVEIS

Assim como com a amostra, as variáveis também precisaram ser alteradas para serem utilizadas na modelagem. A variável sexo (C010) foi transformada numa *dummy* com valor igual a 0 se o indivíduo for homem e 1 se for mulher. A variável C060 – Estado Onde Nasceu, também foi transformada em uma *dummy*, igual a 0 se o indivíduo nasceu no Rio Grande do Sul e 1 caso contrário. Cor também foi transformada em uma *dummy*, com o valor igual a 0 se o indivíduo é branco e 1 caso contrário.

Foi criada a *dummy* “casados” a partir da variável C040 – Posição na Família. Todos os “filho”, “outro parente”, “agregado”, “pensionista”, “empregado doméstico”, “parente do empregado doméstico” e “outro” foi considerado não casado e atribuído o valor 0. Todos os cônjuges foram considerados casados e atribuído o valor 1. Para determinar se os indivíduos considerados chefes de família eram ou não casados, se utilizou o número da família. Se o chefe estava em uma família que também possui conjuge, foi considerado casado, e caso esteja em uma família sem cônjuge foi considerado não casado.

Também foram criados três vetores de *dummies*: um para o ramo de atividade, outro para a ocupação e outro para classe de trabalhador. Os diversos ramos de empresa estão na base de acordo com a Classificação Nacional de Atividades Econômicas (CNAE), e foram agregados em seis grandes grupos, seguindo a Apresentação da Base de Dados do PED. Esses grandes grupos são: indústria de transformação, construção, comércio, serviços, demais setores e sem declaração. Na amostra, não há indivíduos cuja empresa está sem declaração nem empresas do ramo de construção, de modo que esses grupos foram desconsiderados. Assim, o vetor possui três *dummies*: indústria de transformação, comércio e serviços, e demais setores é a variável de referência.

Conforme descrito anteriormente, são utilizados os Grandes Grupos de Ocupação da CBO para analisar as ocupações. O grupo 0 é utilizado como variável de referência. O grupo 6 possui apenas 22 pessoas na amostra, gerando colinearidade na regressão, de modo que essas pessoas foram excluídas e o grupo desconsiderado. Conforme analisado no capítulo 3.4, o nível de agregação das indústrias afeta o resultado encontrado, sendo um nível de maior desagregação preferível, pois auferir resultados mais acurados (FIELDS; WOLFF, 1995). Porém foram escolhidos os dez grandes grupos pela sua praticidade de análise e agregação, já que o próximo subgrupo do CBO é dividido em 47 ocupações, dificultando a formulação do modelo e análise dos resultados.

Por fim, o vetor de *dummies* para classe de trabalhador, adicionada seguindo Oaxaca (1973). Para o mesmo foi utilizada a variável Posição na Ocupação dos Ocupados no Trabalho Principal (Padronizada), ou POS, construída pelo PED. Na amostra obtida, não estão presentes observações de trabalhadores “autônomos para o setor público”, “autônomos para empresa(s)”, “empregadores”, “empregados domésticos diaristas”, “trabalhadores familiares”, “donos de negócio” e aqueles cuja

variável “não se aplica” (valor referentes a 5, 6, 7, 9, 10, 11 e -1 na POS). Dessa forma, o vetor possui apenas seis classes de trabalhadores: assalariados do setor privado com carteira de trabalho assinada, assalariados do setor privado sem carteira de trabalho assinada, assalariado do setor público, assalariado que não sabe a que setor pertence a empresa em que trabalha, empregado doméstico e outras posições na ocupação. Os trabalhadores pertencentes a outras posições na ocupação foram utilizados como variável de referência. Como apenas duas pessoas não sabem a que setor pertencem as empresas em que trabalham, ao estimar o modelo com as mesmas, problemas foram identificados nessa variável, de forma que esses dois indivíduos foram excluídos e a variável desconsiderada.

Além das alterações e criações de *dummies*, também foram realizadas mudanças nas variáveis contínuas. Os valores 10.000.000 e 10.000.001 da variável Q424 – Quanto o Sr.(a) Ganhou pelo Trabalho Realizado no Mês Passado? (Remuneração Líquida no Trabalho Adicional), foram igualados a zero para evitar viés nos dados, já que 10.000.000 na base significa “não se aplica” e 10.000.001 “sem declaração”. Ao total 7.081 trabalhadores da amostra possuíam o valor dessa variável igual a 10.000.000, o que provavelmente indica que não possuem trabalho adicional, e apenas 4 não tinham declaração.

Seguindo a mesma lógica, os valores iguais a 1.000 e 1.001 da variável Q432 – Quantas Horas o Sr.(a) Trabalhou Efetivamente em Cada um dos Seus Trabalhos na Semana Passada? (No Trabalho Adicional), que representam respectivamente “não se aplica” e “sem declaração”, também foram igualados a zero.

Com essas alterações nas variáveis referentes à renda adicional, é possível calcular a renda total por hora de cada trabalhador. Para isso, são somadas as rendas líquidas mensais do trabalho principal e adicional (Q422 e Q424). Esse valor é dividido pelo total de horas trabalhadas na semana (Q431 e Q432) multiplicado por quatro, já que um mês possui quatro semanas. Sendo assim, temos que $rendatotal = \frac{(Q422+Q424)}{(Q431+Q432)*4}$. É importante ressaltar que as horas efetivamente trabalhadas se referem à jornada de trabalho e também às horas além da jornada de trabalho utilizadas em preparação para o trabalho principal ou adicional. São descontadas as horas da semana anterior em que o indivíduo não trabalhou por qualquer motivo, seja ele atestado ou feriado. Esse fato poderia explicar os

indivíduos da base que foram excluídos por terem uma renda do trabalho principal diferente de zero, porém zero horas trabalhadas no trabalho principal.

Educação é tratada como uma constante, de modo que os resultados da variável C102 foram transformados em anos de ensino estimados. O primeiro grau completo foi estimado como representando 8 anos de estudo, o segundo 11 anos e o terceiro 16 anos, e aqueles que nunca frequentaram a escola não tiveram nenhum ano de estudo, ou seja, 0. Foi testado um vetor de *dummies*, com uma *dummy* para cada grau completo (primeiro, segundo e terceiro grau) e “nunca frequentou” como variável de referência. Os resultados encontrados na decomposição em uma e duas partes foram os mesmos dos que os apresentados na próxima sessão, e nenhuma das *dummies* de educação se destacou na análise gráfica das variáveis. Os coeficientes demonstraram o resultado esperado, com os coeficientes de graus superiores sendo maiores. Dessa forma, para simplificar a análise, foi considerada a variável educação com os anos estimados ao invés do vetor de *dummies*.

Por fim, a variável Q40A – Tempo de Emprego (Anos) é usada como *proxy* de experiência. Seguindo a equação minceriana, utilizada em quase todos os modelos apresentados no capítulo 3, é calculada também a experiência ao quadrado.

O modelo que será apresentado não conta com algumas variáveis presentes em outros modelos estudados. Possivelmente a principal delas é a variável filhos, presente nos modelos de Oaxaca (1973), parcialmente presente em Corcoran e Duncan (1979) através dos dias de trabalho faltados em função de questões não relacionadas com a saúde do próprio indivíduo e estudada por Mincer e Polachek (1974), cujos resultados empíricos apontam para a não significância dessa variável. A mesma poderia ter sido estimada de modo similar aos indivíduos casados: aqueles em famílias que possuem uma pessoa cuja posição é “filho”, o chefe e o conjugue seriam considerados como tendo filhos, e caso contrário como não tendo filhos. Porém essa estimação estaria sujeita a diversos erros, como por exemplo o “filho” na verdade ser o chefe de família, o “cônjuge” não ser um dos pais do “filho”. Além disso, na categoria “filho” da PED também são considerados enteados, o que não faria sentido na análise, já que a presença de filhos é relevante quanto ao período de abstinência dos pais no mercado de trabalho (MINCER; POLACHEK, 1974). No entanto, Blinder (1973), Fields e Wolff (1995) e Pereira e Oliveira (2013) também não incluem a presença de filhos como variável em seus modelos.

Outra variável que não é utilizada na análise é uma em relação à saúde do indivíduo, que pode afetar sua produtividade no trabalho. Oaxaca (1973) e Blinder (1973) utilizam variáveis sobre a saúde em seus modelos, e Corcoran e Duncan (1979) controlam por dias faltados por causa de doenças. Informações sobre a saúde dos indivíduos não é disponível no PED, de modo que não foi inclusa. Ainda assim, Mincer e Polachek (1974), Fields e Wolff (1995) e Pereira e Oliveira (2013) também não controlam por essa variável.

Dessa forma, são sumarizadas abaixo as variáveis utilizadas.

Renda total: variável dependente; total líquido recebido em reais pelo indivíduo por hora.

Sexo: *dummy* igual a 1 se a pessoa for mulher e 0 se a pessoa for homem.

Idade: idade em anos do indivíduo.

Estado civil: *dummy* igual a 1 se casado e 0 caso contrário.

Cor: *dummy* igual a 0 se branco e 1 caso contrário.

Imigração: *dummy* igual a 0 se a pessoa nasceu no Rio Grande do Sul e 1 caso contrário

Educação: anos de estudo estimado de cada indivíduo (8 para aqueles com primeiro, 11 para o segundo e 16 para o terceiro grau completo).

Experiência: anos de trabalho de cada indivíduo (termo linear e quadrático).

Ocupação: vetor com oito *dummies* dos grandes grupos ocupacionais do CBO, sendo o grupo 6 excluído e o grupo 0 variável de referência.

Classe de trabalhador: vetor com 5 *dummies* (trabalhador do setor privado com carteira assinada, trabalhador do setor privado sem carteira assinada, trabalhador do setor público e empregado doméstico), sendo outras posições na ocupação a variável de referência.

Ramo de atividade: vetor com quatro *dummies* dos ramos de atividades das empresas onde os indivíduos trabalham (indústria de transformação, comércio, construção e serviços), sendo demais setores a variável de referência.

4.4 MODELAGEM E ANÁLISE DOS RESULTADOS

Tendo analisado as variáveis individualmente no capítulo anterior, é possível construir a função da renda utilizada. ε é o termo de erro, que se supõe ter distribuição normal e média zero.

$$\begin{aligned}
 \text{rendatotal} = & \\
 & \alpha + \beta_1 \text{sexo} + \beta_2 \text{idade} + \beta_3 \text{casados} + \beta_4 \text{cor} + \beta_5 \text{imigrante} \\
 & + \beta_6 \text{educação} + \beta_7 \text{experiência} + \beta_8 \text{experiência}^2 + \beta_9 \text{ocupação} \\
 & + \beta_{10} \text{classe de trabalhador} + \beta_{11} \text{ramo de atividade} + \varepsilon
 \end{aligned}$$

A variável sexo é desconsiderada e são feitas regressões separadas para homens e mulheres, cujos coeficientes são apresentados na tabela 1. Dessa forma, é possível realizar a decomposição de Oaxaca.

O salário médio por hora estimado é R\$ 10,87 para homens e R\$ 10,21 para mulheres, existindo assim uma diferença salarial de R\$ 0,66 por hora. Desse modo, a proporção do rendimento por hora das mulheres em relação aos homens na amostra utilizada é de 93,9%, sendo maior que a encontrada pela FEE (2018), de 88,0% ao considerar a totalidade dos dados da amostra.

Tabela 1 - Coeficientes das Regressões da Função Renda para Homens e Mulheres.

Variável	Regressão para Homens	Regressão para Mulheres
Intercepto	-4,5289198 ***	-9,604763
Idade	0,0893249 ***	0,049121 ***
Casados	-0,0552992	0,617304 **
Cor	-0,6918327 **	-0,262059
Imigração	2,3489452 ***	0,815074
Educação	0,7177032 ***	0,629527 ***
Experiência	0,2005596 ***	0,301702 ***
Experiência ²	-0,0005027	-0,003579 *
Ocupação		
Grupo 1	8,4018960 ***	0,621928
Grupo 2	12,8515288 ***	6,363795
Grupo 3	0,6651974	-3,094860
Grupo 4	-2,9586826 ***	-5,353192
Grupo 5	-2,6897569 **	-6,134166
Grupo 7	-2,3174484 **	-7,230859
Grupo 8	-1,8046719	-6,591192
Grupo 9	-1,2031633	-6,057851
Classe de Trabalhador		
Empregado doméstico	8,6462899 ***	14,672972 **
Privado sem carteira	5,7148635 ***	16,409267 **

Privado com carteira	4,7970512 ***	14,793335 **
Setor público	11,3343410 ***	16,645707 **
Ramo de atividade		
Ind. de transformação	-0,6423769	-0,575748
Comércio	-1,2160445 **	-2,001427
Serviços	-0,9567757	-1,471717
n	3.798	3.309
R²	0.457892	0.411768
R² ajustado	0.454733	0.407830

Fonte: Elaboração própria (2018).

Os asteriscos representam o nível de significância de cada variável em cada uma das regressões, conforme o padrão. Nota-se o fato de mais variáveis serem significativas na regressão masculina do que na feminina. Idade, educação e experiência são altamente significativas para ambas as regressões. Além disso, todas as variáveis *dummy* referente à classe de trabalhador também são significativas para ambas regressões.

É interessante notar que o coeficiente para a variável de estado civil possui sinais contrários para homens e mulheres. Enquanto para homens ele é negativa, ou seja, estar casado diminuiria a renda (mesmo que de maneira diminuta), para mulheres é positivo.

Outro ponto notável são os coeficientes da variável idade: para os homens, o mesmo tem um valor duas vezes maior do que para as mulheres, o que aponta para a existência do fator idade apontado por Blinder (1973), segundo o qual os homens, conforme a idade, conseguem posições e cargos melhores e, conseqüentemente, rendas maiores, e as mulheres, não. Imigração é outra variável cujo coeficiente é maior para os homens. Como o mesmo é positivo, pode significar que um número maior de homens se muda em função do emprego, assim melhorando sua posição, enquanto as mulheres se mudam em função do marido ou da família, tendo que obter um novo emprego na cidade para onde vão (MINCER; POLACHEK, 1974).

A variável cor possui um coeficiente negativo para ambos, conforme esperado, de modo que não brancos recebem menos que brancos de ambos os sexos. Educação e experiência possuem coeficientes similares para os dois sexos, e experiência ao quadrado também, possuindo um sinal negativo conforme indica a literatura.

Os coeficientes dos diferentes grupos de ocupação e das diferentes classes de trabalhadores variam muito para os diferentes sexos, enquanto os de ramo da atividade possuem uma variação menor e são todos negativos. Nos diferentes grupos ocupacionais, todos os coeficientes são maiores para homens, enquanto na classe de trabalhador o contrário ocorre.

Os coeficientes dos grupos ocupacionais podem ser uma evidência de discriminação estatística, sendo as mulheres concentradas em ocupações que remuneram menos. O grupo 1, composto por membros superiores do poder público, dirigentes de organizações de interesse público e de empresas e gerentes (BRASIL, 2017) possui o maior diferencial entre os coeficientes. Cerca de 40% dos cargos na Justiça Comum de primeira Instância no Brasil são ocupados por mulheres. Esse número cai para 29% para cargos de magistrados na Justiça Federal, e vão diminuindo conforme instâncias e cargos superiores são considerados (SOUZA, 2008). No setor privado, 16% dos cargos de CEO no Brasil são ocupados por mulheres, de acordo com a pesquisa realizada pela Grant Thorn (2017). Dessa forma, as evidências apontam para uma menor participação de mulheres nesse grupo, corroborando a hipótese de discriminação estatística.

Com as regressões estimadas é realizada a decomposição de Oaxaca-Blinder. A mesma pode ser realizada de duas maneiras: separando a equação do diferencial em duas partes ou em três (do inglês, decomposição do tipo *twofold* ou *threefold*). Sendo $\Delta\bar{Y}$ o valor médio da diferença da renda, \bar{X}_A o vetor dos valores médios das variáveis explicativas para o grupo A (no caso homens), \bar{X}_B o vetor dos valores médios das variáveis explicativas para o grupo B (no caso mulheres) e $\widehat{\beta}_B$ e $\widehat{\beta}_A$ os coeficientes estimados para cada grupo, se obtém a equação de duas partes abaixo, igual a feita por Oaxaca (1973) explicada na seção 3.2.1. O termo $(\bar{X}_A - \bar{X}_B)\widehat{\beta}_B$ seria a parte explicada da diferença salarial, enquanto $\bar{X}_A(\widehat{\beta}_A - \widehat{\beta}_B)$ seria a parte não explicada, geralmente atribuída à discriminação (HLAVAC, 2018).

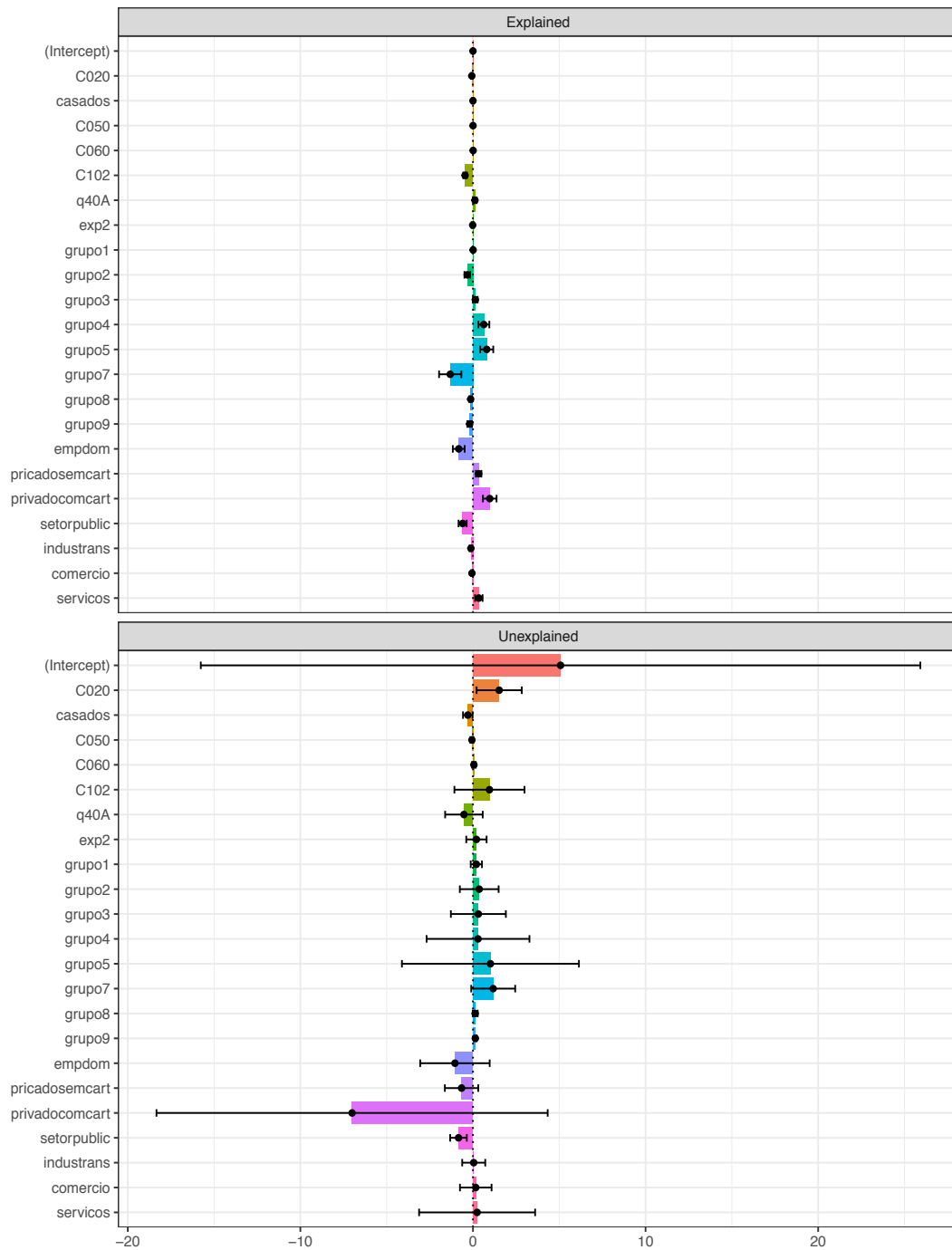
$$\Delta\bar{Y} = (\bar{X}_A - \bar{X}_B)\widehat{\beta}_B + \bar{X}_A(\widehat{\beta}_A - \widehat{\beta}_B)$$

A mesma diferença pode ser decomposta em três partes, conforme a equação abaixo, que é similar ao trabalho de Blinder (1973), também apresentado na seção 3.2.1. O termo $(\bar{X}_A - \bar{X}_B)\widehat{\beta}_B$ representa a diferença devido a diferenças nas

variáveis explicativas (atributos ou *endowments* em inglês), $\overline{X}_B(\widehat{\beta}_A - \widehat{\beta}_B)$ mostra a diferença devido a diferentes coeficientes para os dois grupos, e $(\overline{X}_A - \overline{X}_B)(\widehat{\beta}_A - \widehat{\beta}_B)$ é o termo de interação (HLAVAC, 2018).

$$\Delta \bar{Y} = (\overline{X}_A - \overline{X}_B)\widehat{\beta}_B + \overline{X}_B(\widehat{\beta}_A - \widehat{\beta}_B) + (\overline{X}_A - \overline{X}_B)(\widehat{\beta}_A - \widehat{\beta}_B)$$

Gráfico 10 - Decomposição de Oaxaca-Blinder em Duas Partes



Fonte: Elaboração própria (2018).

Utilizando a amostra e a equação de renda já apresentados, é feita a decomposição de Oaxaca separada em duas e em três partes. A decomposição em duas partes resultou em -0,77 reais por hora para a parte explicada e 1,43 reais por hora para a parte não explicada. Isso significa que, de acordo com suas variáveis explicativas, era de se esperar que as mulheres recebessem R\$ 0,77 por hora a mais que os homens. No entanto, a parte não explicada resulta em um diferencial de R\$ 1,43 por hora em favor dos homens. Essa diferença é geralmente atribuída a discriminação, e nesse caso a totalidade da diferença salarial no mercado de trabalho seria resultado de discriminação.

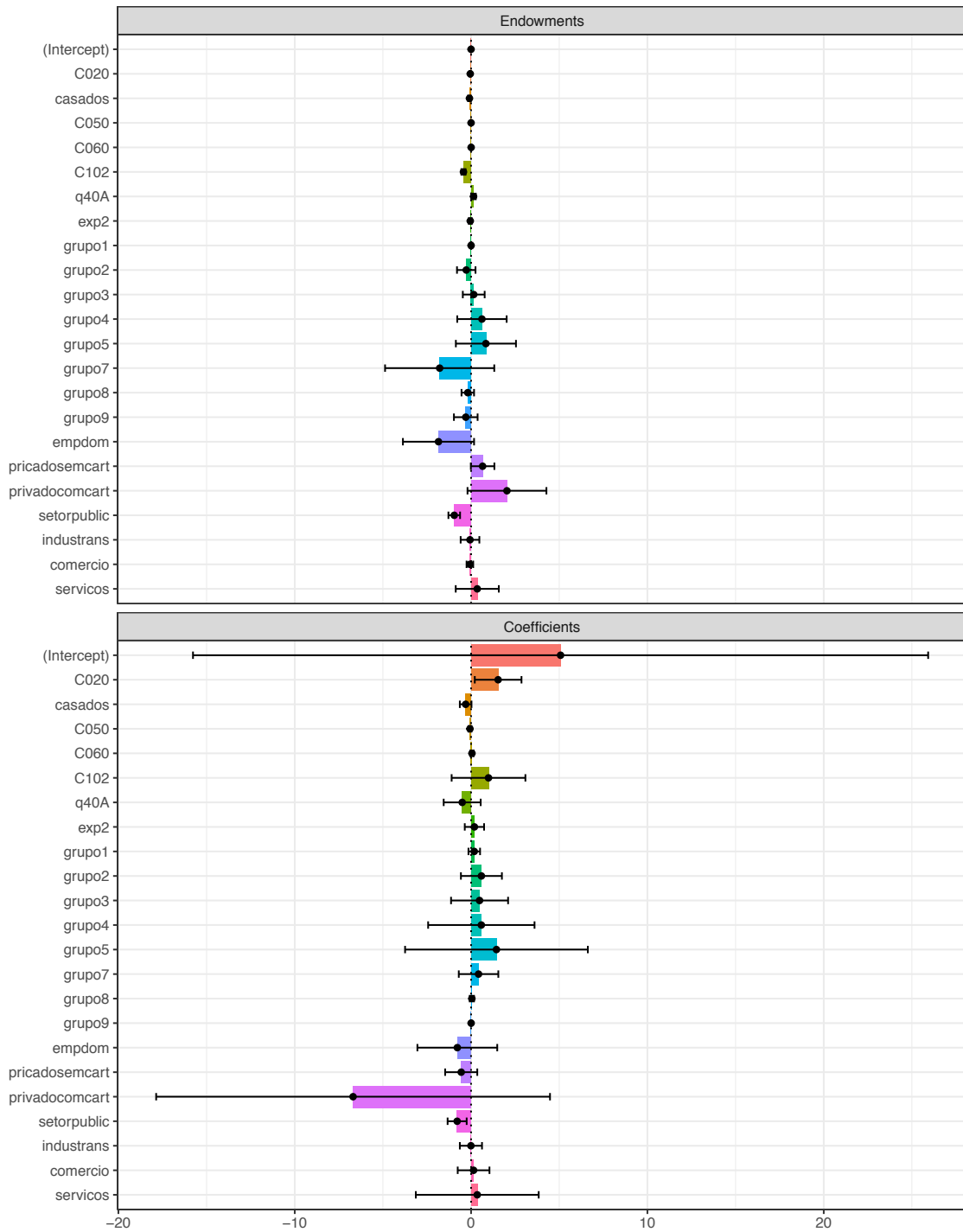
No gráfico 10, as barras representam o impacto de cada variável para o resultado do termo explicado da diferença (*explained*, do inglês) e não explicado (*unexplained*, do inglês), e as linhas mostram o intervalo de confiança de 95%. A linha que divide o gráfico ao meio representa a situação sem diferença salarial. As barras para a direita representam as variáveis que representam uma vantagem para os homens na diferença, e as para a esquerda uma vantagem para as mulheres.

A parte explicada não possui muitas variáveis relevantes, com a maioria das barras se situando próximas ao eixo. Já o contrário ocorre na porção não explicada, onde alguns componentes se destacam. O primeiro deles é o intercepto, que mostra que para qualquer nível das demais variáveis, os homens recebem um salário maior. No entanto, o mesmo pode apresentar grande relevância devido às variáveis de controle escolhidas para os três vetores de *dummies*, cujo impacto é capturado pelo intercepto (HORRACE; OAXACA, 2001). Dessa forma, não é possível afirmar se sua relevância se dá devido às mesmas ou à discriminação. A idade também possui relativa significância na parte não explicada da diferença salarial, apontando novamente para o efeito idade identificado por Blinder (1973).

A *dummy* para trabalhadores do setor privado com carteira assinada também se destaca, porém em favor das mulheres. Os dados da FEE (2018) mostram que, para a totalidade da amostra, essa categoria é onde as mulheres possuem um maior rendimento por hora em relação aos homens, de 91,7%. Além disso, é a única classe de trabalhador cujo rendimento real médio aumentou para as mulheres (em 0,4%) e diminuiu para os homens (em 3,4%), apesar de os homens continuarem recebendo mais do que as mulheres (FEE, 2018). Dessa forma, o trabalho no setor privado com carteira assinada apresenta uma vantagem relativa para as mulheres. Todavia, como a maior parte da amostra pertence a essa classe de trabalhador

(74,3%, de acordo com o gráfico 7), esse resultado pode refletir um problema nos dados decorrente disso.

Gráfico 11 - Decomposição de Oaxaca-Blinder em Três Partes



Fonte: Elaboração própria.

Os resultados da decomposição de Oaxaca em três partes mostram novamente que as mulheres deveriam receber 1,28 reais por hora a mais que os

homens, de acordo com seus atributos (*endowments* em inglês, tradução nossa) como educação e experiência. No entanto de acordo com os coeficientes os homens receberiam 2,38 reais por hora a mais que as mulheres, ou seja, eles recebem mais pelos mesmos atributos. Já o coeficiente de interação mostra uma diferença nos rendimentos de 0,83 reais por hora a favor dos homens.

Em relação aos coeficientes, novamente o intercepto, a idade e a *dummy* de trabalhadores do setor privado com carteira assinada se destacam. A interpretação para esses fenômenos está em consonância com a discutida anteriormente para a decomposição em duas partes. Apesar disso, a *dummy* dos trabalhadores do setor privado com carteira assinada beneficia os homens na questão dos atributos, o que pode ocorrer devido ao seu maior número nessa classe de trabalhador.

A *dummy* do grupo 5 de ocupações também se destaca na análise dos coeficientes para a decomposição em três partes. O grupo 5 engloba os trabalhadores de serviços e vendedores do comércio em lojas e mercados (BRASIL, 2017). De acordo com a FEE (2018), entre 2016 e 2017 houve uma retração de 9,2% de postos de trabalhos no setor de serviços ocupados por mulheres, diminuindo a sua participação no mesmo, o que pode explicar parcialmente a relevância dessa variável para a diferença salarial. Porém no mesmo período, foi constatado um aumento da igualdade de rendimentos nesse setor (FEE, 2018). É importante ressaltar que os dados utilizados se referem apenas ao ano de 2017, de modo que não capturam efeitos temporais.

Analisando as principais variáveis dos atributos (*endowments* em inglês, tradução nossa), as *dummies* do grupo 7 e de empregados domésticos se destacam em favor das mulheres. O grupo 7 representa os trabalhadores da produção de bens e serviços industriais, como trabalhadores da indústria extrativa e da construção civil, trabalhadores da transformação de metais e de compósitos, entre outros (BRASIL, 2017). Entre 2016 e 2017 houve uma diminuição da desigualdade de rendimentos entre homens e mulheres na indústria de transformação. Além disso, a participação das mulheres nesse setor no mesmo período aumentou em 5,5%, enquanto a dos homens diminuiu em 0,6% (FEE, 2018). No entanto, conforme ilustram o gráfico 2 e 3, apenas 8,6% das mulheres da amostra são empregadas nessa ocupação, em contraste com 33,3% dos homens. Ainda assim, esse dado leva a uma outra explicação: como os homens estão mais concentrados nas

ocupações desse grupo, que pagam menos, a variável do mesmo contribui em favor das mulheres quando analisada a diferença salarial.

Em relação ao emprego doméstico, as mulheres dominam o setor, compondo quase a totalidade de empregadas domésticas. Com o aumento do desemprego entre 2016 e 2017, houve também uma maior precarização do trabalho, levando a um aumento de 4,4% de empregadas domésticas (FEE, 2018). No entanto, era de se esperar que a concentração de mulheres nessa atividade, que é considerada de menor qualidade de menor remuneração, contribuísse para a diferença salarial, e não o contrário, como ocorre. Apesar de a renda ficar abaixo da média, entre 2005 e 2011 seu aumento real foi acima dos demais trabalhadores (SOUZA; DOMINGUES, 2014). Esta tendência pode ter permanecido, justificando o benefício que o trabalho doméstico apresenta para as mulheres na questão do diferencial salarial.

O pacote “oaxaca” do software R permite identificar a variável de referência quando são utilizadas *dummies*. Ao fazer isso, o pacote cria o termo $a = \frac{\sum_{j=1}^{k-1} \beta_j}{k}$, onde k é o número de *dummies* e β_j o coeficiente da *dummy* de referência j . O termo a é então adicionado ao coeficiente do intercepto e subtraído dos demais coeficientes (HLAVAC, 2018). No entanto, o pacote não comporta mais de um vetor de *dummies*, de modo que essa função não pode ser utilizada no modelo. O efeito desse controle de sensibilidade foi testado para a equação controlando para um vetor de *dummies* de cada vez, e se concluiu que apenas os erros padrões sofrem alterações quando se usa essa função. Desse modo, o modelo é sensível às variáveis de referência e possui complicações nos erros padrões.

Apesar de a modelagem apresentar os problemas nos erros padrões discutidos, ainda é possível analisar os resultados encontrados, que não parecem ter sido afetados pelas mesmas. Ambas as decomposições feitas apontam para a presença de discriminação na diferença salarial entre homens e mulheres na região metropolitana de Porto Alegre. Seguindo a literatura investigada no capítulo 3, a menor remuneração das mulheres é em parte explicada pela discriminação, seja na parte não explicada na decomposição de duas partes ou na diferença dos coeficientes na decomposição de três partes. As principais variáveis que influem essa diferença a favor dos homens é o intercepto, a idade e a *dummy* dos trabalhadores pertencentes ao grupo 5. Por outro lado, as principais variáveis que atuam em prol das mulheres na diferença são a *dummy* dos trabalhadores do setor

privado com carteira assinada, a *dummy* dos trabalhadores pertencentes ao grupo 7 e a *dummy* de empregados domésticos.

O mesmo modelo foi empregado no relatório publicado pela FEE sobre o mercado de trabalho do Rio Grande do Sul, utilizando dados da Pesquisa Nacional por Amostra de Domicílio (PNAD). As variáveis utilizadas são similares às dos estudos já analisados, porém alguns pontos se destacam. Primeiramente, é utilizado o rendimento mensal, e não por hora como feito no presente trabalho e demais estudos analisados. Além disso, a idade é utilizada como *proxy* de experiência, de modo que também é utilizado seu valor quadrático. O mesmo é feito com a variável “idade em que começou a trabalhar”. Os autores também adicionam a variável estimada “probabilidade de interromper a carreira profissional”, calculada utilizando dados anteriores do PNAD (STEIN; SULZBACH; BARTELS, 2015). Por fim, cabe destacar que utilizam a proporção de mulheres que trabalham naquele tipo de ocupação e atividade (STEIN; SULZBACH; BARTELS, 2015), porém não utilizam *dummies* para os diferentes setores e atividades

Os resultados encontrados apontam para uma diferença salarial de 20,8% a favor dos homens. Além disso, também mostram que as mulheres são mais escolarizadas, assim como os dados apresentados na sessão 4.32 e, em média, dedicam três vezes mais horas para os afazeres domésticos que os homens. As mulheres também se concentram em ocupações específicas e tem uma probabilidade maior de interromper a carreira do que os homens.

Ao construir o modelo, os autores concluem que 13,5% dessa diferença salarial é atribuída às características individuais, e apenas o restante seria decorrência de discriminação. 12,7% seria explicado pelo perfil de emprego, “mulheres tendem a se concentrar em ocupações e atividades que remuneram, em média, menos” (STEIN; SULZBACH; BARTELS, 2015, p. 16). Além disso, os autores argumentam que a menor jornada de trabalho das mulheres também contribuiria para esse diferencial (STEIN; SULZBACH; BARTELS, 2015). Quando a renda média é dividida pela média das horas trabalhadas, a diferença salarial cai para 8% (ou seja, próximo ao nível que seria explicado por discriminação).

5 CONSIDERAÇÕES FINAIS

A diferença salarial entre homens e mulheres é amplamente reconhecida e divulgada. Seu grau e justificativas, por outro lado, além de variarem são questionados e debatidos. Em especial, a presença ou não de discriminação na determinação desses salários é um aspecto que gera discussões. O presente trabalho buscou responder o seguinte questionamento: “como é determinada diferença salarial na região metropolitana de Porto Alegre?”. Para o mesmo, foram desenvolvidos objetivos específicos que visavam a análise dos fenômenos estudados.

O primeiro objetivo específico, desenvolvido no capítulo 2, era uma revisão bibliográfica da teoria econômica da discriminação. Discriminação não é um tema comumente tratado na análise econômica, como aponta Becker (1971). O autor desenvolve um método de agregar a discriminação na análise econômica, desenvolvendo um coeficiente de discriminação que transformaria as preferências pessoais em valores monetários. Além disso cria o coeficiente de discriminação de mercado (BECKER, 1971).

Além dessa interpretação econômica de discriminação, existe também a de discriminação estatística, que é resultado do custo da informação. Dessa forma, alguma característica comum de um grupo é utilizada como *proxy* de sua produtividade (CORCORAN; DUNCAN, 1979). Devido à dificuldade de aplicação dessa definição em modelos salariais mais amplos, sendo sua análise empírica feita mais comumente por experimentos, como realizado por Dickinson e Oaxaca (2009).

Com essa base teórica, é possível realizar o segundo objetivo, de revisar os principais modelos utilizados para analisar a diferença salarial. São apresentados quatro modelos principais: o baseado no capital humano desenvolvido por Mincer e Polachek (1974), a decomposição de Oaxaca-Blinder (1973), o modelo de estimativa de produtividade de Cox e Nye (1989) e o modelo interindustrial, elaborado por Fields e Wolff (1995) e revisado por Horrace e Oaxaca (2001). Com exceção do Cox e Nye (1989), todos os modelos e estudos que utilizam os mesmos analisados encontraram evidências de discriminação de gênero.

Para a análise empírica foi escolhido o modelo de Oaxaca e Blinder (1973), devido ao seu amplo uso, além das variáveis disponíveis. Utilizando os dados do PED, que foram filtrados de modo a adequá-los ao trabalho, é feita a decomposição em duas e três partes, ambas evidenciando a presença de discriminação na diferença salarial de R\$ 0,66 por hora de homens e mulheres na região metropolitana de Porto Alegre.

Uma análise da decomposição mostra que essa diferença se dá devido à discriminação, pois de acordo com seus atributos (a parte explicada da diferença salarial) as mulheres receberiam mais que os homens, no entanto o mercado remunera mais os homens. A idade se mostrou um fator importante em favor dos homens, o que pode significar que os mesmos tendem a ter uma melhoria de posição no cargo durante os anos maior que as mulheres. A variável do grupo 5, dos trabalhadores de serviço, também contribui em favor dos homens para a diferença salarial. Como esse grupo ocupa um maior número de mulheres, é possível concluir que o mesmo tem uma remuneração mais baixa que os demais.

Cabe destacar que algumas variáveis contribuem em favor das mulheres na diferença salarial, embora sua interpretação se mostre mais desafiadora. A variável de trabalhadores do setor privado com carteira assinada é de grande relevância, porém o mesmo pode ocorrer pelo fato de 66,9% das mulheres da amostra pertencerem a esse grupo. Outra variável cuja análise se mostrou de maior complexidade é a de trabalhadores domésticos, que favorece as mulheres na diferença salarial. Isso poderia ser explicado pelo fato de 428 das 441 pessoas da amostra que são empregados domésticos serem mulheres. Outra possível explicação seria as trabalhadoras domésticas receberem mais do que os homens. Por fim, a variável do grupo 7, de trabalhadores da indústria, também favorece as mulheres. Sua análise é análoga à do grupo 5; o grupo 7 emprega 33,3% e apenas 8,6% das mulheres. Por ser um grupo cuja remuneração é baixa, isso afeta a diferença salarial em favor das mulheres.

O estudo possui limitações, sendo a base de dados limitada, assim como o modelo. Além das limitações da base e conseqüentemente do modelo, o mesmo ainda apresenta problemas nos erros padrões devido à não especificação das variáveis de referência nos vetores de *dummies*. Ainda assim, pretende-se mostrar que a discriminação deve ser estudada e considerada em análises do mercado de trabalho.

REFERÊNCIAS

- AIGNER, Dennis J.; CAIN, Glen G.. Statistical theories of discrimination in labor markets. **Industrial and Labor Relations Review**, Ithaca, v. 30, n. 2, p. 175-187, Jan. 1977. Disponível em: <<http://www.jstor.org/stable/2522871>>. Acesso em: 21 mar. 2018.
- ARROW, Kenneth J.. What has economics to say about racial discrimination?, **Journal of Economic Perspective**, Stanford, v. 12, n. 2, p. 91-100, 1998. Disponível em: <<https://goo.gl/Tuk3ZB>>. Acesso em: 20 mar. 2018.
- BECKER, Gary. **The economics of discrimination**. 2nd ed. Chicago: The University of Chicago Press, 1971.
- BLINDER, Alan S.. Wage discrimination: reduced form and structural estimates. **The Journal of Human Resources**, Madison, v. 4, n. 4, p. 436-455, 1973. Disponível em: <<http://www.jstor.org/stable/144855>>. Acesso em: 28 mar. 2018.
- CORCORAN, Mary; DUNCAN Greg J. Work history, labor force attachment, and earnings differences between the races and sexes. **The Journal of Human Resources**, Madison, n. 1, p. 3-20, 1979. Disponível em: < <https://bit.ly/2kPr3yK>>. Acesso em: 11 set. 2017.
- COX, Donald; NYE, John Vincent. Male-Female wage discrimination in nineteenth-century France. **The Journal of Economic History**, Cambridge, v. 49, n. 4, p. 903-920, Dez. 1989. Disponível em: < <https://bit.ly/2JtEHpk>>. Acesso em: 01 jun. 2017.
- DICKINSON, David L.; OAXACA, Ronald L.. Statistical discrimination in labor markets: an experimental analysis. **Southern Economic Journal**, Tucson, v. 76, n. 1, p. 16-31, 2009. Disponível em: <<https://bit.ly/2xO9nNA>>. Acesso em: 11 set. 2017.
- DUNCAN, Otis Dudley. Family investments in human capital: earnings of women: comment. **Journal of Political Economy**, Chicago, n. 2, p. 109-110, Mar. – Abr. 1974. Disponível em: < <https://bit.ly/2Jmk0Za>>. Acesso em: 11 set. 2017.
- FIELDS, Judith; WOLFF, Edward N.. Interindustry wage differentials and the gender wage gap. **Industrial Labor Relations Review**, Ithaca, v. 49, n. 1, p. 105-120, Out. 1995. Disponível em: <<https://bit.ly/2HoITlg>>. Acesso em: 28 mar. 2018.
- FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA - FEE. **A inserção da mulher no mercado de trabalho da região metropolitana de Porto Alegre**. Porto Alegre, mar. 2018.

_____. **Pesquisa de Emprego e Desemprego na Região Metropolitana de Porto Alegre**. Porto Alegre, 2017. Disponível em <<http://ped.fee.tche.br/>>. Acesso em: 21 mar. 2018.

GOLDIN, Claudia. The changing economic role of women: a quantitative approach. **Journal of Interdisciplinary History**, Cambridge, v. 13, n. 4, p. 707-733, 1983. Disponível em <<https://bit.ly/2JB9Wiw>>. Acesso em: 07 dez. 2017.

GRANT THORN. **Women in business**: novas perspectivas sobre risco e oportunidade. [S.l.], 2017. Disponível em <<https://bit.ly/2sKLWiq>>. Acesso em: 15 maio 2018.

HECKMAN, James J.. Sample selection bias as a specification error. **Econometrica**, New York, v. 47, n. 1, p. 153-161, Jan. 1979. Disponível em <<https://bit.ly/2JouAip>>. Acesso em: 13 mar. 2018.

HLAVAC, Marek. **oaxaca**: Blinder-Oaxaca Decomposition in R. R package version 0.1.4. [S.l.], 2018. Disponível em <<https://CRAN.R-project.org/package=oaxaca>>. Acesso em: 02 mar. 2018.

HORRACE, William C.; OAXACA, Ronald L. Inter-Industry wage differentials and the gender wage gap: an identification problem. **Industrial and Labor Relations Review**, Ithaca, v. 54, n. 3, p. 611-618, Abr. 2001. Disponível em <<https://bit.ly/2Lp4a0y>>. Acesso em: 11 set. 2017.

MESSINA, Julián; SILVA, Joana. **Wage inequality In Latin America**: understanding the past to prepare for the future. Washington: The World Bank, 2017. Disponível em <<https://bit.ly/2sz2gDG>>. Acesso em: 04 jun. 2017.

MINCER, Jacob; POLACHEK, Solomon. Family investments in human capital: earnings of women. **Journal of Political Economy**, Chicago, v. 82, n. 2, p. 76-108, Mar. – Abr. 1974. Disponível em <<https://bit.ly/2LqreN>>. Acesso em: 11 set. 2017.

BRASIL. Ministério do Trabalho. **Classificação Brasileira de Ocupações – CBO**. 2017. Brasília. Disponível em: <<https://bit.ly/1pO9ADe>>. Acesso em: 03 abr. 2018.

OAXACA, Ronald. Male-female wage differentials in urban labor markets. **International Economic Review**, Amherst, v. 14, n. 2, p. 693-709, Out. 1973. Disponível em: <<https://bit.ly/2HsnNSS>>. Acesso em: 11 set. 2017.

OAXACA, Ronald. Some observations on the economics of women's liberation. **Challenge**, Oxfordshire, v. 19, n. 3, p. 29-33, Jul. – Ago. 1976. Disponível em: <<https://bit.ly/2slvGyl>>. Acesso em: 11 set. 2017.

ORGANIZAÇÃO DA NAÇÕES UNIDAS NO BRASIL - ONUBR. **Indígenas, negros e mulheres são mais afetados por pobreza e desemprego no Brasil, diz CEPAL**. [S.l.], set. 2016. Disponível em <<https://bit.ly/2f53DkV>>. Acesso em: 12 dez. 2017.

PEREIRA, Rafael Mesquita; OLIVEIRA, Cristiano Aguiar. **Os diferenciais de rendimento por gênero no Rio Grande Do Sul**: uma aplicação do modelo de

Heckman e da decomposição de Oaxaca-Blinder. Trabalho apresentado no 26. Encontro de Economia da Região Sul, Curitiba, 2013. Disponível em: <<http://www.anpec.org.br/sul/2013/>>. Acesso em: 07 dez. 2017.

PHELPS, Edmund S.. The statistical theory of racism and sexism. **The American Economic Review**, Pittsburg, v. 62, n. 4, p. 659-661, Set. 1972. Disponível em: <<https://bit.ly/2sBc6VB>>. Acesso em: 22 mar. 2018.

SAFFIOTI, Heleieth. Papéis sociais atribuídos às diferentes categorias de sexo. In: SAFFIOTI, Heleieth. **O poder do macho**. São Paulo: Moderna, 2001. p. 8-19.

SOUSA, Ana Júlia da Silva de. Participação da mulher nos espaços de poder no Brasil: atuação feminina no executivo, legislativo e judiciário. In: **Âmbito Jurídico**, Rio Grande, XIV, n. 91, ago 2011. Disponível em: <<https://bit.ly/2xPCPme>>. Acesso em: 15 maio 2018.

SOUZA, Kênia Barreto de; DOMINGUES, Edson Paulo. Mudanças no mercado de serviços domésticos: uma análise da evolução dos salários no período 2006-2011. **Economia Aplicada**. São Paulo. v. 18, n. 2, p. 319-346, 2014.

STANFORS, Maria; LEUNIG, Tim; ERIKSSON, Björn; KARLSSON, Tobias. Gender, productivity, and the nature of work and pay: evidence from the late nineteenth-century tobacco industry. **The Economic History Review**, Oxford. v. 67, n. 1, p. 48-65, 2014.

STEIN, Guilherme; SULZBACH, Vanessa Neumann; BARTELS, Mariana. **Relatório sobre o mercado de trabalho do Rio Grande do Sul — 2001-13**. Porto Alegre: FEE, 2015.