

Diferenciais de Rendimentos entre Homens e Mulheres no Brasil revisitado: explorando o “Teto de Vidro”[#]

Renato Vale Santos *

Eduardo Pontual Ribeiro **

Resumo

Diferencias de rendimentos negativos médios para as mulheres em relação aos homens, mesmo controlando atributos produtivos, são fatos estilizados na literatura nacional e internacional. Por outro lado, o comportamento destes diferenciais em toda a distribuição de rendimentos é quase inexplorado. Este trabalho identifica o diferencial de rendimentos entre homens e mulheres, controlando por atributos produtivos para o Brasil em todos os níveis de rendimentos, ao invés de focar o rendimento médio. Com isto podemos estudar a existência do chamado “*glass ceiling*” ou *teto de vidro* nos rendimentos femininos de modo condicional. Para isto empregamos uma metodologia econométrica, desenvolvida por Machado e Mata (2005, J. Applied Econometrics) que permite a estimação de toda uma distribuição contrafactual com menos hipóteses do que os trabalhos existentes na literatura brasileira. Os resultados mostram que existem diferenciais não explicados pelos atributos produtivos, sendo bastante heterogêneos. As mulheres beneficiam-se destes diferenciais de tratamento nos rendimentos mais baixos e são prejudicadas nos rendimentos mais altos da distribuição. Desta forma pode-se argumentar pela existência do fenômeno de *glass ceiling*, mas também de um possível *glass floors* para o Brasil.

Palavras Chave: diferenciais de rendimento, gênero, discriminação, regressão nos quantis

JEL: J16, J31, C15

Abstract

Mean earnings differentials against women, even controlling for productive characteristics, are a stylized fact in the empirical labor economics literature. On the other hand, the shape of these differentials along the wage distribution are almost unknown. This paper measures the extension of unexplained earnings differentials between men and women across the earnings distribution, using Machado and Mata (J. Appl. Econometrics, 2005) methodology. This allows us to evaluate the glass ceiling hypothesis for Brazil, controlling for observed productive characteristics. Our results suggest that the unexplained earnings differentials are heterogeneous. These differentials benefit women at lower earnings, contrary to the mean and higher earnings differentials that benefit men. One could argue that there is a glass ceiling in Brazil, but also a kind of “glass floor”.

Key words: earnings differentials, gender, discrimination, quantile regression

JEL: J16, J31, C15

COMENTÁRIOS BEM VINDOS

[#] O trabalho é baseado na dissertação de mestrado do primeiro autor. Agradecemos o apoio computacional e esclarecimentos de Juliana Guimarães, José A. Machado e Flávio Ziegelmann, assim como comentários de Giacomo Balbinotto e Naércio Menezes.

* Centro Universitário Newton Paiva/MG e PPGE/UFRGS

** IE/UFRJ, PPGE/UFRGS e Pesquisador do CNPq. Contato: eribeiro@ie.ufrj.br

INTRODUÇÃO

A existência de diferenciais nos rendimentos médios entre homens e mulheres não explicados por características produtivas é fato conhecido na literatura internacional (Altonji e Blank, 1999, por exemplo) e nacional (Corseuil et al., 2002 ou Loureiro, 2003, entre outros). A relevância destes diferenciais fez com que um dos Objetivos de Desenvolvimento do Milênio, liderado pela ONU com apoio de quase todos os países do mundo, inclusive o Brasil, fosse exatamente “Promover a igualdade entre os sexos”.

A mensuração usa métodos de regressão para separar a parcela dos diferenciais que seria devido às diferenças de características produtivas pessoais e outra parcela que seria devido a diferenças na remuneração destas características, parcela muitas vezes chamada de discriminação, por ser não explicada por fatores observáveis. Devido aos métodos de regressão empregados na análise, esta se restringe à avaliação de diferenciais de rendimentos *médios*.

Todavia há a crença de que os diferenciais aumentam ao longo das distribuições de rendimentos, sendo este efeito apelidado em vários países de *glass ceiling* ou teto de vidro. Este *teto* seria uma barreira invisível que impediria o acesso de mulheres a postos e ocupações de melhores rendimentos. Recentemente, com o lançamento da *Síntese de Indicadores Sociais* de 2005 pelo IBGE em abril de 2005, foi noticiado na imprensa que os diferenciais (não controlados) de rendimentos por gênero aumentam ao longo dos decis de renda e aumentam por grupo de anos de estudo. Naquela publicação pode-se ver que enquanto as mulheres, em média, recebem 84% do rendimento dos homens, este diferencial aumenta de 80,8% para pessoas com até 4 anos de estudo, para 61,6% para pessoas com 12 anos ou mais de estudo.

Focando apenas o emprego formal, baseado na RAIS/MTE de 2002, o diferencial se mantém. Segundo o nível de escolaridade, os rendimentos das mulheres analfabetas superam em 0,8% dos trabalhadores nessa mesma classificação no setor formal. Nos demais níveis, o salário médio feminino é sempre inferior ao masculino. Mesmo nos níveis superior incompleto e completo, pela Figura 1, a remuneração média feminina corresponde a 61% e 57,9%, respectivamente, dos salários percebidos pelos homens.

Não há muitos trabalhos na literatura econômica internacional que tratem da questão do teto de vidro. Albrecht, Bjorklund e Vroman (2001) e delaRica, Dolado e Llorens (2005) são exceções. Neles fica clara a dificuldade de tratar da análise da persistência de um diferencial crescente com a renda (o teto de vidro) nos rendimentos controlados pelas características pessoais, pois é necessário estimar um modelo de regressão para caracterizar toda a distribuição condicional dos rendimentos, e não somente a sua média (e variância). Os autores usam métodos de regressão quantílica, introduzidos por Koenker e Bassett (1978) para fazer a estimação da distribuição condicional de rendimentos.

Ao mesmo tempo, este trabalho tenta verificar o que aconteceria com os diferenciais de rendimentos *não condicionais* entre homens e mulheres se o tratamento diferenciado de características pessoais fosse eliminado. Para isto, empregamos um método de construção de distribuições contrafactuais desenvolvido por Mata e Machado

(2005), que, de acordo com Melly (2005), é superior em vários aspectos aos métodos de simulações contrafactuais baseados em modelos de regressão por mínimos quadrados (modelos previsão de médias condicionais), como por exemplo em Juhn, et al. (1996) e Blau e Khan (1997) e, no Brasil, Soares (2002) e Barros, Corseuil e Leite (2000).

O artigo está dividido em quatro partes, além desta introdução. Na próxima seção, iremos descrever melhor os dados empregados, baseados na PNAD de 1999 e uma melhor caracterização do *glass ceiling* no Brasil. Adiantando os resultados, o aumento do diferencial de rendimentos não é monotônico. Na segunda seção, especificamos o modelo empírico a ser estimado, assim como a metodologia de estimação por regressão nos quantis (*quantile regression*) e o método de Mata e Machado(2005) para simulações contra-factuais. A terceira seção traz os resultados e a última os comentários finais.

1. Diferenciais de Rendimentos e Discriminação no mercado de trabalho no Brasil

Para exemplificar a situação do diferenciais de rendimentos entre homens e mulheres em profundidade, usamos tabulações a partir dos microdados da Pesquisa Nacional por Amostragem de Domicílios (PNAD) de 1999. Os valores são atualizados para valores de dezembro de 2003. Foram selecionados assalariados com rendimento positivo com idade entre 25 e 55 anos, período em que a participação no mercado de trabalho é máxima tanto para homens como para mulheres. Os rendimentos de todos os trabalhos foram padronizados pelo número de horas trabalhadas para evitar que a jornada de trabalho influenciasse na avaliação dos diferenciais de rendimentos. As estatísticas descritivas do logaritmo dos rendimentos-hora de homens e mulheres obtidos da amostra da PNAD podem ser observados na Tabela 1. São apresentadas medidas de tendência central, alguns percentis, duas medidas de dispersão (desvio padrão e coeficiente de variação) e uma medida de desigualdade distributiva (o índice de Theil).

Analisando os principais percentis e a média, percebemos que os rendimentos-hora dos homens são, em todos os casos, maiores que das mulheres. Este pode ser um indício, que a distribuição de rendimentos dos homens domina estocasticamente em primeira ordem a distribuição dos rendimentos das mulheres. O diferencial dos rendimentos a favor dos homens é menor para o primeiro decil, onde os homens recebem 3,5% a mais que as mulheres. O diferencial aumenta até a mediana das distribuições, quando alcança o valor máximo de 14,3%. A partir daí, o diferencial cai no percentil 75% e volta a aumentar no último decil, quando chega a 12,5% do rendimento hora das mulheres no mesmo decil de sua distribuição.

Tabela 1 – Estatísticas Descritivas ln (rendimentos-hora)

Estatísticas	Homens	Mulheres	Diferencial (%)
Média	0,9019	0,8153	9,0%
Percentil 10	-0,2231	-0,2578	3,5%
Percentil 25	0,2231	0,1252	10,3%
Percentil 50	0,7621	0,6286	14,3%
Percentil 75	1,4271	1,3863	4,2%
Percentil 90	2,2381	2,1203	12,5%
Desvio Padrão	0,9491	0,9319	

Theil	0,891	0,852
-------	-------	-------

Fonte: cálculo do autor baseado em dados da PNAD.
Diferencial refere-se a diferença do rendimento dos homens em relação aos das mulheres.

Para uma melhor visualização dos diferenciais de rendimentos, apresentamos na Figura (2) a diferença dos rendimentos entre homens e mulheres ao longo das distribuições acumuladas. Vemos, assim como na tabela acima, que o diferencial de rendimentos aumenta até a mediana, tende a cair até o decil 8 e cresce rapidamente a partir daí. Como todos os diferenciais de percentis são positivos, podemos afirmar que a distribuição de rendimentos dos homens domina em primeira ordem a das mulheres e que verifica-se o teto de vidro nos diferenciais não controlados.

Se por um lado, não verificamos diferenciais negativos para os homens em relação às mulheres em nenhum ponto da distribuição, por outro o comportamento não monotônico destes diferenciais é interessante. Na verdade o teto de vidro pode ser mais forte do que se imagina, pois as mulheres nos percentis 60 a 80 (onde o diferencial é decrescente) podem ter maiores qualificações que homens com rendimentos similares. Numa análise controlada por características produtivas, o gráfico deste diferencial pode ser monotônico e mais inclinado.

Uma questão interessante associada aos diferenciais de rendimentos entre homens e mulheres é o impacto da discriminação na desigualdade de renda. O coeficiente de Theil calculado mostra que a distribuição dos rendimentos dos homens é menos equitativa que das mulheres. Todavia, segundo o relatório da PNUD(2001), ambos apresentam coeficientes elevados para os padrões internacionais. Na seção seguinte, com base nos coeficientes estimados pela regressão quantílica poderemos calcular a curva de Lorenz, e visualizar como esta desigualdade se comporta para todos os quantis. Podemos visualizar melhor esta dispersão relativa entre homens e mulheres pelo *Box Plot* como mostra a figura (3).

A caixa no *Box Plot* é construída pelas linhas referentes aos valores do percentil 25, da mediana e do percentil 75. Os pontos na parte superior do gráfico representam os valores discrepantes (outliers). Como podemos observar, as mulheres apresentam concentração dos dados maior na parte inferior do gráfico (2º quartil), mas se observarmos o intervalo interquartílico, as mulheres apresentam dispersão maior que os homens (a caixa das mulheres é um pouco maior que dos homens). Além disso, há presença de grande quantidade de outliers para as duas distribuições.

1.1. Explicando e mensurando os diferenciais de rendimentos

Na literatura econômica, estes diferenciais foram bastante estudados e associados à questão da discriminação. Borjas (1996) determina a existência de discriminação econômica no mercado de trabalho, quando diferenças de rendimentos e de oportunidades de emprego ocorrem entre trabalhadores com habilidades iguais no mesmo emprego. Isto é, trabalhadores com mesmo nível educacional, treinamento, experiência e produtividade têm seus rendimentos diferenciados devido à sua raça, sexo, religião, orientação sexual, nacionalidade e outras características irrelevantes para o exercício de seu trabalho. Muitas teorias tentam justificar economicamente a existência e persistência desta discriminação, para empresas inseridas em ambientes competitivos.

Neste trabalho não fazemos uma resenha da volumosa literatura sobre discriminação, dada a existência de vários trabalhos que fazem esta síntese, como Loureiro (2003), Corseuil (2002), Altonji e Blank (1999) e Borjas (1996), limitando-nos a indicar as idéias mais importantes na área.

Em sua revisão da literatura, Loureiro (2003) considera a discriminação no mercado de trabalho de um modo mais amplo do que apenas diferenças de rendimentos, classificando a discriminação em quatro tipos: discriminação salarial, discriminação de emprego, discriminação de trabalho ou ocupacional e discriminação ao acesso de capital humano.

- Discriminação salarial: onde os trabalhadores discriminados recebem salários menores do que os outros trabalhadores, fazendo o mesmo trabalho.
- Discriminação de emprego: em que momentos de baixa oferta de empregos, os trabalhadores discriminados são colocados em desvantagem, sendo, portanto mais atingidos pelo desemprego.
- Discriminação de trabalho ou ocupacional: ocorre quando os trabalhadores discriminados são impossibilitados de assumir certas ocupações, tendo a mesma capacidade que os trabalhadores não discriminados, de executar tais trabalhos.
- Discriminação decorrente de oportunidade desigual para se obter capital humano: ocorre quando os trabalhadores discriminados encontram dificuldades ou são impossibilitados de aumentar sua produtividade, mediante educação ou treinamento no trabalho.

Na literatura econômica, os três primeiros tipos são denominados *postmarket discrimination*, devido ao fato da discriminação ocorrer quando os trabalhadores já se encontram no mercado de trabalho. O último é denominado *premarket discrimination*, porque ocorre antes do indivíduo entrar no mercado de trabalho. Barros e Mendonça (1996) trazem uma taxonomia similar, ao identificar fatores de mercado e extra-mercado que geram desigualdade de rendimentos.

Três fontes gerais de discriminação no mercado de trabalho foram teorizadas e cada uma sugere um modelo de como a discriminação é implementada e quais são as suas conseqüências (Ehremberg e Smith, 2000). A primeira fonte de discriminação é o preconceito pessoal, em que empregadores, empregados ou consumidores não aceitam o

convívio no ambiente de trabalho com funcionários pertencentes a um grupo de características específicas (sexo, raça, etc). Esta classe de modelos teve início com o trabalho seminal de Becker (1957). A segunda fonte é o preconceito estatístico. Em um ambiente de informação incompleta, empregadores usam características pessoais produtivas e, principalmente, não produtivas, para projetar nos indivíduos algumas características *médias* notadas no grupo. Desta forma, trabalhadores com produtividade acima da média recebem rendimentos abaixo de seu potencial. Estes modelos foram iniciados por Phelps (1972), Aigner e Cain (1977) e Spence (1973). E, por último, existem modelos de discriminação nos quais o desejo e uso do poder de monopólio são a fonte de discriminação, como em Piore (1970) e Bulow e Summers (1986).

Em todos os modelos a discriminação ocorre quando trabalhadores não são remunerados de modo "justo" ou seja, de acordo com suas produtividades. Desta forma a caracterização da discriminação se dá quando não é possível justificar, a partir de medidas de produtividade pessoais, os diferenciais de rendimentos.

O modo mais popular de mensuração da discriminação é a decomposição Oaxaca-Blinder (Oaxaca, 1973, e Blinder, 1973). Supondo dois grupos de trabalhadores, por exemplo, homens (h) e mulheres (m), a média dos salários dos homens é dada por $\overline{w_h|X}$ e a média de rendimentos das mulheres é dada por $\overline{w_m|X}$. O diferencial de rendimentos médios $\Delta\overline{w|X}$ é dada por

$$\Delta\overline{w|X} = \overline{w_h|X} - \overline{w_m|X}.$$

Este diferencial não implica em discriminação visto que a discriminação ocorreria se os rendimentos de pessoas com a mesma produtividade fosse diferenciado. Isto pode ser mensurado definindo os rendimentos condicionais médios à características produtivas dos homens como $\overline{w_h|X}$ e a média dos salários condicionais das mulheres como $\overline{w_m|X}$. A medida de discriminação é dada pela diferença dos salários médios condicionais, ou seja, padronizados pelas características produtivas das pessoas.

$$\Delta\overline{w|X} = \overline{w_h|X} - \overline{w_m|X} \quad (2)$$

A contribuição de Oaxaca e Blinder foi identificar como relacionar $\Delta\overline{w}$ e $\Delta\overline{w|X}$. Isto é possível através de regressão por mínimos quadrados e suas propriedades.

Para uma exposição mais didática, suponhamos que apenas uma variável (escolaridade), denotada por S , afeta os rendimentos. A função de ganho para cada um dos dois grupos pode ser escrita como:

$$w_h = \alpha_h + \beta_h s_h + \varepsilon_h \quad (3)$$

$$w_m = \alpha_m + \beta_m s_m + \varepsilon_m$$

Pelas propriedades de MQO, os resíduos de cada regressão têm média zero e a regressão avaliada na média da(s) variáveis explicativa(s) tem como valor previsto dos rendimentos sua média não condicional, ou seja,

$$\bar{w}_h = \alpha_h + \hat{\beta}_h \bar{S}_h \quad (4)$$

$$\bar{w}_m = \alpha_m + \hat{\beta}_m \bar{S}_m$$

Onde \bar{S}_h e \bar{S}_m são respectivamente as médias de escolaridade dos homens e mulheres. O coeficiente $\hat{\beta}_h$ mostra quanto aumenta o ganho dos homens para um aumento de um ano de escolaridade, e o coeficiente $\hat{\beta}_m$, de forma análoga, apresenta a mesma estatística para as mulheres. O modelo de regressão implica que o diferencial de salários brutos pode ser escrito como:

$$\Delta \bar{w} = \bar{w}_h - \bar{w}_m = \alpha_h + \hat{\beta}_h \bar{S}_h - \alpha_m - \hat{\beta}_m \bar{S}_m \quad (5)$$

Podemos decompor o diferencial de salários brutos $\Delta \bar{w}$ em duas partes. Uma parte que determina o diferencial de salários devido às diferenças de características (habilidades) entre os indivíduos, e outra parte que determina o diferencial devido à discriminação. Esta decomposição é conhecida como decomposição de Oaxaca (1973) ou Oaxaca-Blinder. Para o nosso caso, basta somar e subtrair uma média artificial, ou contrafactual de w , dada pelo produto dos coeficientes da regressão das mulheres e a dotação média de atributos dos homens, $\alpha_m + \hat{\beta}_h \bar{S}_m$, à equação (1.14) e, através de manipulações algébricas, temos:

$$\Delta \bar{W} = \underbrace{(\alpha_h - \alpha_m)}_{\text{Diferencial devido a discriminação}} + \underbrace{(\hat{\beta}_h - \hat{\beta}_m) \bar{S}_m}_{\text{Diferencial devido a diferentes habilidades}} + \hat{\beta}_h (\bar{S}_h - \bar{S}_m) \quad (6)$$

Assim, a equação (6) mostra que o diferencial de salário bruto consiste em duas partes. A interpretação é bem simples: se homens e mulheres possuem a mesma escolaridade média (ou $\bar{S}_h - \bar{S}_m = 0$), o segundo termo da equação (6) será zero. Portanto, uma parte do diferencial de salários brutos surgirá quando os dois grupos possuírem diferentes habilidades.

Se, por acaso, para um empregador a utilidade da escolaridade de um homem é maior que a utilidade da escolaridade de uma mulher $\hat{\beta}_h > \hat{\beta}_m$, ou ainda, se o empregador paga mais a homens do que a mulheres para qualquer nível de escolaridade (o intercepto da função de rendimento é maior para homens do que para mulheres; $\alpha_h > \alpha_m$), o primeiro termo da equação será positivo. Dessa forma, a diferença de salários surgirá pelo tratamento diferencial dos homens em relação às mulheres, demonstrando uma situação de discriminação.

A validade da decomposição de Oaxaca para medir discriminação depende se temos controle para todas as dimensões em que as habilidades dos dois grupos diferem e são correlacionadas com os controles incluídos na regressão. Se há algumas características de habilidade que afetam os ganhos que são correlacionadas com as características incluídas na regressão, mas que nós deixamos fora do modelo teremos medidas incorretas da discriminação no mercado de trabalho por viés e/ou inconsistência dos coeficientes (sejam angulares ou lineares).

Uma limitação da decomposição de Oaxaca é a de que se baseia em modelos de regressão, que geram médias condicionais. Portanto temos informações referentes apenas aos valores de rendimentos médios, o que pode ser insuficiente para entender a

discriminação no mercado de trabalho. Não sabemos onde a discriminação é maior ou menor na escala de salários. Podemos ter discriminação muita elevada com relação às mulheres para níveis de salários muito elevados e, em contrapartida, pouca ou nenhuma discriminação nos níveis salariais mais baixos e este tipo de informação não é fornecido através da média condicional. Em suma, sua utilização com métodos de análise de regressão na média (ou até na mediana) não permite avaliar a questão do teto de vidro. Neste texto empregamos um método de regressão (quantílica) para superar esta limitação.

1.2. Discriminação e diferenciais de rendimentos por gênero no Brasil

No Brasil, o problema da discriminação e/ou diferenciais de salários no mercado de trabalho foi estudado por diversos autores, vários já resenhados nos *surveys* de Coresuil (2002) e Loureiro (2003). Dentre eles destacamos Barros e Ramos (1992), Barros et alii (1996), Ometto et alii (1999), Carvalho (2004) e referencias ali contidas, Giubertti e Menezes-Filho (2005), Cavalcanti e Guimarães (1999) e Guimarães (2001).

A grande maioria destes trabalhos utilizou a técnica econométrica usual de MQO para trabalhar com os dados e realizar as estimativas. Utilizando a técnica de regressão quantílica, encontramos na literatura os trabalhos recentes de Cavalcanti e Guimarães (1999) e Guimarães (2001), sendo que a última, adota a metodologia de Machado e Mata (2004).

A mensagem comum de todos os trabalhos é que os diferenciais de rendimentos *médios* não explicados pelas características são grandes e significativos contra as mulheres, existindo em diferentes cortes (rural/urbano, formal/informal, entre ocupações), períodos e regiões. Barros e Ramos (1992) argumentam que, se for levado em conta a idade e a educação, o grau de discriminação será maior do que se apenas forem estimadas as diferenças de salários entre homens e mulheres sem levar em conta estas variáveis.

Dentre os trabalhos encontrados que utilizam regressão quantílica para tratar diferencial de rendimentos e/ou discriminação por gênero. Destacamos o trabalho pioneiro, apesar de limitado regionalmente, de Cavalcanti e Guimarães (1999), que utiliza regressão quantílica para calcular o coeficiente de discriminação de Oaxaca. O trabalho foi realizado com base nos dados da PNAD para as Regiões Metropolitanas de Recife e São Paulo em 1989. Os resultados obtidos mostram que os retornos a educação são positivos para todo os quantis, e que são crescentes ao longo da escala de rendimentos tanto para os homens como para as mulheres. Mas, os retornos à educação crescem mais rapidamente para os homens que para as mulheres. A consequência é que investimentos em educação não implicam, diretamente em melhor distribuição de renda.

Em relação à discriminação, Cavalcanti e Guimarães (1999) constataram a presença de discriminação em relação às mulheres, sendo esta crescente dos níveis iniciais de rendimento para os mais elevados. Fenômeno este atribuída em parte a menor dispersão da distribuição dos rendimentos das mulheres em relação aos homens. Indicando que o aumento proporcional do número de trabalhadores do sexo feminino em relação aos homens, podem contribuir com a equalização da distribuição da renda per capita.

Todavia, a limitação do trabalho é que os diferenciais *condicionais* são crescentes, não havendo a ligação entre rendimentos condicionais e não condicionais. Já Guimarães (2001) utiliza a metodologia de Machado e Mata (2004) para estudar a mudança na distribuição de rendimentos para o Brasil, comparando os anos de 1986 e 1995. Embora não principalmente focado na questão do gênero, pois não estimava regressões separadas para homens e mulheres, o texto traz a primeira aplicação do método de Machado e Mata para o Brasil, construindo distribuições de rendimentos contrafactuais de rendimentos com coeficientes de 1995 e características de 1986.

1.3. Estimativas para o *glass ceiling*

Na literatura econômica sobre diferenciais de rendimento, como em Cavalcanti e Guimarães (1999) e Guimarães (2001), muitas vezes encontramos que o diferencial de rendimentos entre homens e mulheres é crescente. Especificamente, este diferencial apresenta uma aceleração para a parte final da função de distribuição. Este fenômeno é conhecido na literatura econômica como *glass ceiling*. Teoricamente, representaria uma barreira no crescimento nos rendimentos mais elevados das mulheres. Este é um tópico de pesquisa recente e não encontramos nenhum estudo para o Brasil. Para o mercado de trabalho de outros países destacamos três trabalhos que abordam o tem utilizando duas técnicas distintas.

Albrecht, Bjorklund e Vroman (2001) utilizando a técnica de regressão quantílica e a metodologia de Machado e Mata (2004), estudaram o fenômeno para os dados da Suécia e dos Estados Unidos. No caso americano, apesar dos diferenciais de rendimentos serem elevados, os autores não verificaram a presença de *glass ceiling*. Para Suécia, verificaram a presença em vários anos da década de noventa. Um resultado interessante é que Albrecht, Bjorklund e Vroman (2001) realizam o teste para os trabalhadores imigrantes e não-imigrantes da Suécia e não encontraram *glass ceiling*, o que pode se ser um indício de que o fenômeno ocorre especificamente para diferenças de rendimentos por gênero.

delaRica, Dolado e Llorens (2005) também utilizam a técnica de regressão quantílica e a metodologia de Machado e Mata (2004) para estimar os retornos para diferentes características dos trabalhadores na Espanha em 1999. Para comparar com os resultados da Espanha, os autores estimaram os diferenciais de rendimentos de Portugal, Itália, França e Dinamarca. Em todos os países foi observada a presença de *glass ceiling*, mas o diferencial de rendimentos ao longo da função de distribuição da Espanha se assemelha ao verificado para Grécia e Itália. Nestes países, o diferencial de rendimentos em favor dos homens é decrescente ao longo da distribuição de renda, aumentando apenas no final da distribuição. A este efeito os autores denominaram como *glass floors* ou melhor, “*glass ceilings at the floor level*”.

Interessante notar que quando os autores separam a amostra em homens e mulheres com escolaridade alta (universitária) e média/baixa (ensino fundamental e médio), o *glass ceiling* é claro para a escolaridade alta e o *glass floors* fica evidente para a escolaridade média/baixa. O argumento teórico dos autores é que empregadores podem exercer com maior força a discriminação para as mulheres de baixo capital humano, pois o valor presente da maior rotatividade das mulheres em relação aos salários baixos justificaria menores rendimentos em relação aos homens.

2. Metodologia empírica - regressão nos quantis

Mínimos quadrados são o método mais utilizado para realizar estimativas ou fazer previsões denomina-se mínimos quadrados (MQ). Esta ferramenta popularizou-se pela sua facilidade de implementação e por constar na maior parte dos softwares econométricos. MQ é utilizada para estudar as relações entre variáveis, onde desejamos prever uma resposta média da variável dependente, Y , condicionada aos valores assumidos pelas covariáveis, X . E para que média nos forneça uma informação precisa da distribuição de Y condicional a X , sua distribuição deve ser simétrica.

Além da facilidade de implementação computacional, outros fatores ocasionaram o sucesso dos MQ. Entre eles, suas propriedades consideradas ideais como possuir um estimador de mínima variância entre os estimadores não-viesados para os casos em que uma função de distribuição da variável dependente é Normal (Gaussiana). Entretanto, de acordo com o teorema Gauss-Markov, as propriedades do estimador de MQO são ótimas apenas na classe de estimadores lineares, quando os erros não são Normais. Em adição, MQ Não é um estimador robusto, pois sofre grande influência de valores extremos da variável dependente, como explica Koenker (2005).

A técnica de regressão quantílica desenvolvida por Koenker e Basset (1978) além de não necessitar da suposição básica de mínimos quadrados de erros homocedásticos, ela é mais informativa, pois permite estimar toda a distribuição condicional da variável dependente, através dos quantis da distribuição. Podemos obter uma regressão para cada quantil de interesse em contrapartida a MQO que nos fornece apenas a média. Portanto, regressão quantílica nos propicia um estimador mais robusto quando a distribuição dos erros for heterocedástica e/ou não-Gaussiana e fornece uma informação mais completa da distribuição de Y condicionada a X . A Figura (4) mostra uma distribuição onde os erros são heterocedásticos, e como regressão quantílica pode ser mais informativa que MQO.

2.1. Aspectos técnicos de regressão quantílica

A técnica de regressão quantílica foi apresentada inicialmente no trabalho de Koenker e Basset (1978), e pode ser vista como uma generalização do modelo de regressão de Mínimos Desvios Absolutos (MDA), onde podemos estimar a mediana da distribuição de Y condicionada ao valor de suas covariáveis.

Koenker e Basset (1978) apresentam sua técnica definindo a função quantil. Desta forma, dado qualquer valor real da variável aleatória, Y , esta pode ser caracterizada pela sua função de distribuição, dada por:

$$F(x) = P(Y \leq y) \quad (7)$$

onde para qualquer $0 < \tau < 1$, temos a função quantil:

$$F^{-1}(\tau) = Q(\tau) = \inf\{y : F(y) \geq \tau\} \quad (8)$$

em que τ é chamado o τ -ésimo quantil de X . E conseqüentemente $F^{-1}(1/2)$ representa a mediana.

Os parâmetros estimados por regressão quantílica são obtidos da solução de um problema de minimização. O problema consiste em encontrar \hat{y} que minimize o erro

esperado, onde definimos a perda ou erro pela seguinte função linear (função check):

$$\rho_{\tau}(u) = u(\tau - I(u < 0)) \quad (9)$$

A τ -ésima função quantil condicional pode ser representada como $Q_y(\tau | x) = x'\beta(\tau)$, e o vetor de parâmetros $\hat{\beta}(\tau)$ pode ser obtido resolvendo

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - x_i'\beta) \quad (10)$$

Assim, a função ρ_u multiplica os resíduos por $(\tau - 1)$ se eles forem negativos e por τ caso contrário, possibilitando que eles sejam tratados assimetricamente. De acordo com Koenker e Bassett (1978), o problema de regressão quantílica da equação (10) ainda pode ser reformulado como um problema de programação linear, permitindo sua estimação de modo eficiente.

E relação à inferência, no caso em que os erros de um modelo de regressão linear são independentes e identicamente distribuídos (*iid*), estimar a média ou a mediana condicional não propicia diferenças significativas que justifiquem o uso de regressão quantílica em detrimento da estimação clássica por MQO. Segundo Koenker (2005) o modelo clássico de regressão linear com erros *iid* produz uma forma particularmente simples para a distribuição limite do estimador de regressão quantílica $\hat{\beta}(\tau)$. Para o caso de erros não *iid* a matriz de covariância limite, $\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$, toma a forma de Huber (1967), isto é,

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \rightarrow N(0, H_n^{-1} J_n H_n^{-1})$$

onde, $J_n(\tau) = \tau(1 - \tau)n^{-1} \sum_{i=1}^n x_i x_i'$ e $H_n(\tau) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n x_i x_i' f_i(\xi_i(\tau))$.

O termo $f_i(\xi_i(\tau))$ denota a densidade condicional da variável resposta, y_i , avaliada até o τ -ésimo quantil condicional. Observe que se $f_i(\xi_i(\tau))$ não depende de x_i , ou seja, é constante, temos a matriz de variância-covariância para o caso *iid*, $N(0, \tau(1 - \tau)/s(\tau)^2 Q)$, onde $Q = \text{plim } n^{-1} X'X$ e $s(\tau)$ é chamada *sparsity*. Com a especificação da matriz de variância-covariância dos erros, testes de hipótese podem ser facilmente desenvolvidos. Um modo alternativo muito popular e consistente para estimação da matriz de variância-covariância para estimativas por regressão nos quantis é o método do *bootstrap* (Koenker, 2005, Buchinsky, 1998).

O uso de regressão nos quantis é recomendado para a análise de diferenciais de rendimentos, em que a variável dependente é o logaritmo dos rendimentos-hora, por uma propriedade importante, demonstrada por Koenker e Bassett (1978), a *equivariância a transformações monotônicas*. Esta propriedade implica que para $h(\cdot)$ uma função não-decrescente em \mathfrak{R} , então para alguma variável aleatória Y , $Q_{h(Y)}(\tau) = h(Q_Y(\tau))$, isto é, os quantis da variável aleatória transformada $h(Y)$ são simplesmente os quantis transformados da variável original Y . Obviamente a média não possui esta propriedade, ou seja, $Eh(Y) \neq h(E(Y))$, exceto quando $h(\cdot)$ seja uma função linear ou em determinadas situações como sabemos pela Desigualdade de Jensen. A condição (2.13) segue imediatamente do fato elementar que para algum $h(\cdot)$ monotônico, $P(Y < y) = P(h(Y) < h(y))$.

Encontramos freqüentemente por exemplo, $h^{-1}(x'\hat{\beta})$ sendo usado em tais circunstâncias no lugar de $E(y | x)$, e $\exp(x'\hat{\beta})$ quando o modelo tenha sido definido como $\log(y) = x'\beta$. Mas isto gera estimativas viesadas de $E(y | x)$. Em nosso caso específico, isto mostra que os diferenciais absolutos de rendimentos entre homens e mulheres calculados através de regressões mincerianas podem estar superestimados.

2.2. Obtendo distribuições contrafactuais com regressão quantílica: a metodologia de Machado e Mata (2005)

Como visto acima, o método mais popular para identificar diferenciais não explicados, associados à discriminação, nos diferenciais de rendimentos relacionado rendimentos condicionais e não condicionais, através da análise da média das distribuições condicional e incondicional. Para estudar o teto de vidro, necessitamos um método que relacione *toda* a distribuição de rendimentos não condicional com a distribuição condicional. Infelizmente esta é uma tarefa difícil.

Os *percentis* condicionais condicionais, quando avaliados no valor médio das explicativas não gera percentis não condicionais dos rendimentos. Assim, é necessário fazer a decomposição simulando, de alguma forma, toda a distribuição de rendimentos. Esta é a contribuição de Machado e Mata (2005). Há outros métodos para estimar distribuições marginais a partir de distribuições marginais, construindo a assim, distribuições contra-factuais da variável de interesse. Dois candidatos são o método não paramétrico de reponderação de densidades de DiNardo, Fortin e Lemieux (1996) e métodos baseados em regressão, como Juhn, Murphy e Pierce (1997) e Barros et al. (2000). Todavia, os dois métodos alternativos apresentam algumas limitações. No caso de DFL, é complexo fazer a análise com variáveis contínuas ou mais de uma explicativa. Já no caso e JMP, o modelo implícito é mais restrito que o de MM, como é demonstrado por Melly (2005). Assim iremos empregar o método de MM.

O algoritmo de Machado e Mata é baseado na idéia da decomposição de Oaxaca (1973). Convém neste ponto fazermos a associação entre Oaxaca e Machado e Mata, destacando os aspectos estatísticos que fundamentam as duas teorias para a melhor compreensão do processo. Como vimos acima, o método de Oaxaca é baseado no resultado de que $E(y) = E_x[E(y | x)]$. Estatisticamente temos que:

$$\begin{aligned}
 E_x[E(y | x)] &= \int_x E(y | x) f(x) dx = \int_x \left[\int_y y f(y | x) dy \right] f(x) dx \\
 &= \int_x \int_y y \frac{f(x, y)}{f(x)} f(x) dy dx = \int_x \int_y y f(x, y) dx dy \\
 &= \int_y y \left[\int_x f(x, y) dx \right] dy = \int_y y f(y) dy \\
 &= E(y)
 \end{aligned}$$

Cada valor esperado condicionado pode ser obtido através deste processo conhecido em estatística de forma que esses valores esperados são ponderados pela distribuição marginal de x através do operador E_x . O processo consiste no desenvolvimento da alternativa de decomposição da densidade conjunta, sem usar a média diretamente. Para obter um valor esperado não condicional, podemos utilizar a lei das expectativas iteradas, tomando-se valores esperados condicionados por uma das variáveis e depois tomar valores esperados sobre este resultado obtido.

Analogamente podemos associar esta idéia com o algoritmo de MM, que será apresentado em detalhe abaixo. Partindo da definição de regressão quantílica segundo a equação (2.21), podemos escrever $y_\tau | x = F_{y|x}^{-1}(\tau | x) = x\beta_\tau$.

Se τ é gerado aleatoriamente de uma distribuição uniforme, obtemos, $y/x \sim f(y/x)$, para $X = x_0$ usado em $\{x_0 \hat{\beta}_{\tau_j}\}_{j=1}^n$. Para passar de $f(y/x)$ para $f(y)$, temos de integrar para todo x de forma que o exercício empírico consiste em implementar o resultado básico de estatística:

$$\begin{aligned} \int_x f(y|x)f(x)dx &= \int_x \frac{f(x,y)}{f(x)} f(x)dx \\ &= \int_x f(x,y)dx = f(y) \end{aligned}$$

A contribuição de MM é demonstrar que isto é possível fazendo uma amostra aleatória de $X = x_j, j = 1, \dots, n$ em $\{x_j \hat{\beta}_{\tau_j}\}_{j=1}^n$. Com isto, obtemos a densidade marginal dos rendimentos baseado na distribuição condicional estimada. Para gerar uma distribuição contra-factual, valores específicos de X podem ser escolhidos de acordo com o interesse dos pesquisadores.

Finalmente após obtermos as densidades contrafactuais desejadas, podemos decompor toda a mudança nas distribuições de rendimentos para os gêneros em uma parte atribuída aos coeficientes, (β s), outra para as covariáveis (X s) e a última para os resíduos. Além disso, os impactos das covariáveis e dos coeficientes podem ser decompostos para cada um dos atributos individuais considerados no modelo.

Seja $f(w(G))$ a estimativa da densidade marginal de w (log dos rendimentos hora) para o gênero G obtido da amostra, isto é, a densidade empírica, e $f^*(w(G))$ a estimativa da densidade de w em G baseado na amostra gerada $\{w_i^*(G)\}$, isto é a densidade marginal artificial estimada pelo modelo usando novas covariáveis. Estendendo esta notação para as distribuições contrafactuais nos podemos ter: $f^*(w(m); x(h))$ como a densidade que teria prevalecido para as mulheres caso todas as covariáveis tivessem sido distribuídas com a dos homens e as características remuneradas como nas mulheres. Também, $f^*(w(m); z(h))$ seria a densidade dos rendimentos que prevaleceria para as mulheres caso apenas uma covariável, z , tivesse sido distribuída como a dos homens.

Seja $\alpha(\cdot)$ uma estatística descritiva usual como média, quantis ou uma escala de

medida, podemos decompor as mudanças de $f(w(h))$ para $f(w(m))$ em:

$$\begin{aligned} \alpha(f(w(m)) - f(w(h))) &= \alpha(f(w(m)) + f^*(w(m);x(h)) - f^*(w(m);x(h)) - f(w(h))) = \\ &= \alpha[f(w(m)) - f^*(w(m);x(h))] + \alpha[f^*(w(m);x(h)) - f(w(h))] + \text{resíduo} \\ &\quad \text{covariáveis} \qquad \qquad \qquad \text{coeficientes} \end{aligned}$$

O resíduo na equação acima, vem da não linearidade da estatística $\alpha(\cdot)$, isto é, em geral, $\alpha(a+b) \neq \alpha(a) + \alpha(b)$. Comparando a decomposição acima com a decomposição de Oaxaca apresentada acima, naquela é utilizada $\alpha(\cdot)$ como a média e sua linearidade garante que não haja termo residual, pelas propriedades de MQO.

Esta decomposição nos dará diferentes medidas para a contribuição dos coeficientes, as covariáveis e uma parte não explicada (resíduos) para as mudanças da distribuição condicional dos rendimentos observada sobre os gêneros.

3. MODELO ESTIMADO E RESULTADOS EMPÍRICOS

Como dito acima, a base de dados usada foi retirada da Pesquisa Nacional por Amostras de Domicílios (PNAD) de 1999 realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para o Brasil. A equação a ser estimada consiste num modelo *minceriano* de determinação de rendimentos descrito da seguinte forma:

$$\ln w = \alpha + \beta_1 age + \beta_2 age^2 + \beta_3 S_1 + \beta_4 S_2 + \beta_5 S_3 + \beta_6 S_4 + \beta_7 S_5 + \varepsilon \quad (11)$$

onde *age* representa a variável idade, *S* a variável escolaridade e *w* os rendimentos totais. A amostra é limitada a adultos na faixa etária de 25 a 55 anos. A escolha desta faixa etária foi feita com o objetivo de capturarmos as características dos trabalhadores que tivessem maior probabilidade de participação no mercado de trabalho, evitando assim, incluir em nossa amostra, aqueles trabalhadores que estivessem dividindo seu tempo entre trabalho e estudo e os aposentados que estivessem trabalhando para complementar sua renda, para evitar problemas de viés de seleção e tornar os resultados compatíveis com a maior parte da literatura¹.

Os dados originais da PNAD (1999) de rendimentos estão expressos em reais (R\$) mensais. Para evitar a influência da jornada de trabalho, a análise emprega a padronização em rendimento-hora, visto que o número de horas semanais trabalhadas encontra-se disponível na PNAD.

A variável escolaridade (*S*) foi introduzida no modelo da equação de rendimentos através de variáveis *dummy* segundo Ramos e Vieira (1996), para que

¹ Pode-se argumentar pela necessidade de métodos de correção de viés por seleção endógena, dada a menor participação feminina no mercado de trabalho. Neste trabalho não seguimos esta linha pois o método de decomposição de Mata e Machado não é claro, com termos de correção por seleção. Em adição, apesar da popularidade de métodos como Heckit para regressão na média, não há muitas aplicações para correção de viés de seleção usando regressão quantílica. De qualquer forma nossa análise pode ser corretamente interpretada como condicional à amostra de mulheres trabalhadoras. Assim como delaRica, Dolado e Florens (2005), em nenhum momento especulamos sobre o efeito de uma maior participação feminina nos diferenciais de rendimentos.

possamos ver os diferentes efeitos pra cada nível escolar. Desta forma não restringimos os retornos a educação a serem lineares.² Observando que a variável excluída de escolaridade foi nenhuma ou menos de um ano de escolaridade, temos S_1 = de um a três anos de escolaridade; S_2 = de quatro a sete anos de escolaridade; S_3 = de oito a dez anos de escolaridade; S_4 = de onze a quatorze anos de escolaridade; S_5 = acima de quinze anos de escolaridade.

Foram estimadas duas equações, uma para os homens e outra para as mulheres, para que possamos ter os coeficientes estimados para os cálculos das funções densidades e implementação do algoritmo de Machado e Mata (2004). As estimações das regressões quantílicas, bem como das estatísticas descritivas, foram ambas feitas através do software R versão 2.2.0.

Primeiramente, analisamos os coeficientes estimados das regressões quantílicas para o vetor de taus (τ) gerados aleatoriamente segundo o algoritmo de MM. Os resultados são apresentados nas figuras 5, 6 e 7. Os gráficos mostram à esquerda o retorno de cada covariável para os diferentes quantis da distribuição condicional dos rendimentos das mulheres, ao centro, os retornos das estimativas dos homens e a direita a mudança nestes retornos entre os homens e as mulheres. As estimativas são apresentadas através de uma linha cheia, enquanto o intervalo de confiança de 95% é representado pelas linhas pontilhadas. A linha horizontal representa a estimativa por MQO.

A idade foi incluída no modelo com os termos lineares e quadráticos. Portanto seu impacto no logaritmo dos rendimentos-hora foi avaliado usando o efeito marginal, seguindo MM.³ Podemos observar que a variável idade influencia diferentemente o log dos rendimentos-hora para homens e mulheres. Nas mulheres, o retorno à idade é praticamente crescente ao longo da escala do quantis, enquanto para os homens, ele passa a decrescer após o percentil 70%. Comparativamente, observando o gráfico que representa a diferença dos retornos da idade, os homens possuem retornos maiores que as mulheres na maior parte da escala de quantis. Apenas no início e no final da distribuição dos rendimentos quando o retorno da idade dos homens passa a ser decrescente, esta variável tem um retorno maior para as mulheres. É interessante observar como regressão quantílica é mais informativa que MQO. No caso da Idade, se observássemos apenas a média condicional, teríamos o valor da média para as mulheres próximo da mediana e a média dos homens próximo do quantil 0,40. Sendo que o valor do retorno médio dos homens (0,194) é maior que das mulheres (0,167). A conclusão óbvia, é que, a idade dos homens tem um impacto maior no rendimento médio que no caso das mulheres. Mas, pela análise da distribuição dos rendimentos ao longo dos quantis, observamos comportamentos distintos no início e final da distribuição.

Os retornos à escolaridade estão subdivididos em seis níveis, onde o intercepto representa nenhuma ou menos de um ano de escolaridade, e o nível mais elevado representando quinze anos ou mais de escolaridade. No primeiro nível de escolaridade (intercepto), verificamos que os retornos são muito baixos, para homens e mulheres, na maior parte dos quantis e similares aos retornos médios. Apenas para os quantis mais

² Além de Ramos e Soares, Gonzaga e Reis (1999) e outros indicam que os retornos a educação no Brasil são não-lineares.

³ A derivada da equação de rendimentos em relação à idade é dado por : $\frac{\partial w_i}{\partial idade} = \beta_{idade} + 2 * \beta_{idade}$.

elevados (aproximadamente 0,95) está variável apresenta retornos maiores. Em síntese podemos constatar que os rendimentos dos não escolarizados são menores que R\$ 2,00 a hora em quase toda a distribuição.

Na Figura 6, observamos mais dois níveis de escolaridade. Para facilitar a comparação, os gráficos para homens e mulheres possuem a mesma escala em todas as figuras. Os três primeiros gráficos mostram o impacto da variável de um a três anos de escolaridade. Inicialmente observamos a diferença no retorno desta variável entre homens e mulheres. Para as mulheres, os retornos oscilam em torno da média condicional, enquanto para os homens os retornos são crescentes. Já para os homens, os retornos são relativamente bem maiores que para as mulheres. Se olharmos os homens cujos rendimentos estão nos quantis iniciais, o impacto deste nível de escolaridade é semelhante às mulheres cujos rendimentos estão nos quantis mais elevados. O terceiro gráfico retrata bem a diferença no impacto que esta variável tem nos rendimentos de homens e mulheres. A diferença é crescente ao longo da escala de rendimentos, e podemos verificar novamente a possibilidade de *glass ceiling*, embora este pareça mais forte nos rendimentos condicionalmente mais baixos.

Para rendimentos menores, dentro dos níveis de escolaridade, os valores para homens e mulheres são similares. À medida que seguimos na distribuição de rendimentos, o crescimento salarial para os homens é bem mais acentuado do que para as mulheres. Com isso, embora para ambos os gêneros os coeficientes se estabilizem a partir da mediana, o patamar para os homens é bastante superior ao das mulheres. Em adição, no quantis mais elevados, os diferenciais de rendimentos para as mulheres param de crescer, enquanto os homens continuam tendo um impacto crescente.

Os três gráficos restantes da Figura 6 são relativos ao nível de quatro a sete anos de escolaridade. O comportamento desta variável é semelhante ao verificado para o nível de um a três anos de escolaridade. Para os menores níveis de rendimentos (quantis mais baixos) os retornos são similares para homens e mulheres. Estes retornos crescem, mas de modo bastante suave para mulheres e de modo bastante acentuado para homens. A partir do quantil 30 os retornos se estabilizam em torno do valor médio para as mulheres e em um patamar um pouco acima da média para os homens. A partir do percentil 40, os retornos dos homens são cerca de duas vezes maiores do que os das mulheres.

Em suma, três pontos podem ser enfatizados: a menor desigualdade de rendimentos, dentro dos níveis de escolaridades para mulheres, em relação aos homens, a similaridade de rendimentos para salários baixos entre homens e mulheres, dentro do nível de escolaridade, e o pequeno ganho de rendimentos para mulheres ao longo dos decis da distribuição feminina, quando comparado com a distribuição dos homens. O *glass ceiling* aparece a partir do percentil 80, exceto nos níveis de escolaridade mais baixos.

Na Figura 7, visualizamos os três níveis de escolaridade restantes, ou seja, ensino médio, superior incompleto e superior completo. Os três primeiros gráficos referem-se ao nível de 8 a 10 anos de escolaridade. Os três gráficos no meio da figura são relativos à escolaridade de 11 a 14 anos e os três últimos a 15 ou mais anos de escolaridade.

Estes três níveis de escolaridade repetem o comportamento encontrado nos níveis de escolaridade até aqui analisados. Os homens apresentam retornos à escolaridade maior que as mulheres em todos os percentis da distribuição acumulada. Os retornos são bastante diferenciados ao longo da distribuição, sendo menores que a média para quantis inferiores e maiores que a média para quantis superiores. Ou seja, a estimativa de MQO subestima o retorno para altos rendimentos e superestima para baixos rendimentos, dentro do nível de escolaridade. Por outro lado, a diferença de coeficientes é relativamente constante a partir do decil 20 até o decil 80 e muito similar à diferença obtida por MQO. A heterogeneidade de rendimentos parece similar entre homens e mulheres sendo sistematicamente contra as mulheres, com aumento dos diferenciais a partir do decil 80, sugerindo, mais uma vez, a presença de dificuldades de alcançar maiores salários para as mulheres.

A importância de se dividir a escolaridade em níveis, é podermos analisar seus diferentes impactos na distribuição de rendimentos. Cavalcanti e Guimarães (1999) encontraram retornos positivos para educação em todos os quantis. Podemos observar, que na maior parte dos níveis de escolaridade, também encontramos retornos positivos, com exceção dos trabalhadores com nenhuma ou menos de um ano de escolaridade.

Verificamos que homens e mulheres apresentam comportamentos diferentes quanto ao impacto da educação na distribuição dos rendimentos. Esta diferença de comportamento implica que a escolaridade das mulheres contribui para a menor dispersão da distribuição de rendimentos. Este resultado é consistente com o encontrado por Cavalcanti e Guimarães (1999) e Guimarães (2001) para o Brasil, Machado e Mata (2004) para Portugal, Chamberlain (1994) e Buchinsky (1994) para os Estados Unidos, Fitzenberger e Kurz (1997) para Alemanha, González e Miles (2001) para o Uruguai e Nielsen e Rosholson (2001) para a Zâmbia. Ou seja, o impacto da escolaridade e gênero no Brasil apresenta uma característica comum à encontrada em outros países do mundo.

Duas implicações podem ser derivadas do resultado acima. Primeiro, uma redução dos diferenciais de rendimentos devido a retornos à escolaridade entre homens e mulheres pode ter o efeito de aumentar a desigualdade de renda, dado a identificação de retornos mais heterogêneos para homens dentro dos níveis de escolaridade. Mas, se a inserção feminina aumentar com a distribuição de retornos das mulheres, podemos inferir, assim como MM e Guimarães (2001), uma diminuição da dispersão de rendimentos.

Segundo, os retornos crescentes à escolaridade associados a uma heterogeneidade não decrescente dentro dos níveis de escolaridade sugerem que não necessariamente maior educação implica em maior equidade na distribuição da renda, pois um ano adicional de escolaridade têm impactos diferentes sobre os rendimentos para homens e mulheres com o mesmo nível de educação.

Há vários argumentos na literatura para racionalizar estes resultados. Como afirmam Cavalcanti e Guimarães (1999) e outros, talvez uma explicação para estes fenômenos seja a qualidade da escolaridade como recursos importantes para se explicar à heterogeneidade nos retornos da educação. Todavia esta explicação não justificaria a heterogeneidade diferenciada para homens e mulheres, exceto no caso de uma discriminação por gênero já na escola. Há outro argumento associado ao perfil ocupacional de homens e mulheres. As ocupações de homens permitiriam uma maior

heterogeneidade de rendimentos, por terem rendimentos mais altos, em média, dentro das ocupações, do que as mulheres e as ocupações tipicamente masculinas apresentam rendimentos mais altos que as tipicamente femininas, em média.⁴

À medida que a escolaridade aumenta, a heterogeneidade da qualidade da educação – entendida como uma causa da heterogeneidade de rendimentos – parece aumentar, tanto pra homens como para mulheres. Quanto mais especializadas as tarefas no mercado de trabalho, maior a necessidade de habilidades e maior o escopo para diferenças de qualidade na escolaridade, gerando trabalhadores mais heterogêneos.

Pela construção da equação de rendimentos escolhida, conseguimos verificar em todos os níveis de escolaridade uma diferença crescente entre o retorno de escolaridade de homens e mulheres. Nos quantis mais elevados encontramos as maiores diferenças da distribuição de rendimentos, corroborando com a suposição de que ocorrem menores probabilidades de *ascensão* profissional (o chamado *glass ceiling*) que impedem que mulheres com rendimentos mais elevados alcancem o mesmo nível de rendimento dos homens.

3.1 ESTIMAÇÃO CONTRAFACTUAL: ANÁLISE GRÁFICA

Na seção anterior vimos que as diferenças de rendimentos entre homens e mulheres condicional ao nível de escolaridade, gerados pelo tratamento desigual de homens e mulheres (os retornos à escolaridade). Nesta seção iremos identificar o papel destas diferenças de retornos nas diferenças não condicionais de rendimentos por gênero. Nesta seção incluímos os efeitos dos diferenciais de características para explicar os rendimentos observados.

Como descrito na seção de metodologia, podemos decompor as mudanças na distribuição de rendimentos entre os gêneros, em mudanças atribuídas às características observáveis dos trabalhadores, ou atributos (Xs), e mudanças atribuídas aos retornos destas características, ou os coeficientes (βs), e uma parte relativa aos resíduos.

Na figura 8 apresentamos as densidades empíricas obtidas de homens e mulheres ao longo da escala de rendimentos e a densidade contrafactual das mulheres com a função de distribuição de retornos dos homens. Podemos observar que as mulheres possuem uma concentração maior no início da distribuição em relação aos homens. Há uma maior proporção de mulheres que recebem baixos rendimentos, em relação aos homens. Os homens apresentam uma concentração maior para os níveis de rendimentos centrais e mais elevados, exceção para o log dos rendimentos-hora entre três e quatro, onde as mulheres novamente apresentam uma concentração maior. Há uma queda repentina na concentração de mulheres em relação aos homens na cauda direita da distribuição, a partir do log-rendimento igual a 4. Isto pode ser associado ao efeito *glass ceiling*.

Analisando a densidade contrafactual, constatamos que existem ganhos para todas as mulheres caso elas recebessem segundo a distribuição de rendimentos dos homens, pelo deslocamento expressivo da densidade contrafactual para direita. A

⁴ Apesar de ser comum na literatura, a segunda parte da explicação não é corroborado por Araújo e Ribeiro (2002) e Barros et al. (1997) pois os autores identificam que o diferencial médio de rendimentos por gênero não pode ser explicado pelas diferenças de rendimentos médios nas ocupações.

distribuição contrafactual corrobora a identificação do efeito *glass ceiling*. Quando as mulheres recebem retornos aos seus atributos como os homens, a queda da densidade de mulheres com rendimentos acima de 4 é eliminada, se aproximando à distribuição dos homens. A vantagem relativa média das mulheres em atributos em relação aos homens no Brasil, fato estilizado da literatura, implica que a distribuição contra-factual das mulheres está à direita, em quase todos os pontos, do que a distribuição de rendimentos dos homens.

Para estudar o efeito da discriminação na desigualdade empregamos o índice L de Theil parece ser mais apropriado para o nosso estudo de desigualdade. Não apenas devido a sua maior sensibilidade a mudanças nos extremos da distribuição, mas também porque pode ser convenientemente decomposto em componentes que dizem respeito às desigualdades entre-grupos. O índice de desigualdade L de Theil, varia entre 0 e 1. Sendo que quanto maior o índice maior a desigualdade.

Na Tabela 2 apresentamos os índices L de Theil calculados para os dois grupos de gênero, o L de Theil total e o L de Theil entre-grupos. O índice entre-grupos é obtido residualmente pela diferença entre o Theil total e a soma ponderada dos índices dos grupos. Calculamos os respectivos índices com os dados originais de rendimentos e fizemos um exercício contrafactual. Na última coluna da Tabela 2 apresentamos os índices desta simulação, que consiste em calcular o índice L de Theil para as mulheres cujos rendimentos foram obtidos utilizando as covariáveis das mulheres e os coeficientes dos homens.

Tabela 2 – Índice de Theil-L

Índice Theil-L	Dados Originais	Simulação Contrafactual
Theil-L total	0,872	0,894
Theil-L entre-grupos	0,001	0,008
Theil-L Homens	0,891	0,891
Theil-L Mulheres	0,852	0,883

Para os dados originais, temos os homens com desigualdade maior que as mulheres. Observando as Figuras 5 a 7 vemos que a maior variabilidade dos coeficientes dos homens pode justificar a maior desigualdade em relação às mulheres. Quando fazemos a simulação contrafactual e comparamos com os dados originais observamos que o grupo das mulheres teve seu índice de desigualdade elevado. Este resultado é devido à maior variabilidade dos coeficientes dos homens em relação às mulheres, como vemos nas figuras 5 a 7. Conseqüentemente o índice total e o índice entre grupos também aumentaram. Parece haver um certo conflito entre os objetivos sociais de redução de desigualdade e/ou discriminação entre gêneros e a redução da desigualdade na economia como um todo.

A Figura 9 apresenta graficamente a decomposição da diferença das densidades em diferenças de características e diferenças de retornos (associados possivelmente à discriminação). Como em DFL, estas mudanças são obtidas desenhando a diferença entre cada densidade contrafactual e as densidades marginais estimadas para homens ou mulheres. Uma diferença positiva indica que há maior proporção de mulheres naquele nível de rendimentos.

O primeiro gráfico da Figura 9 nos fornece informações das mudanças nas

densidades entre mulheres e homens: $f^*(w(m)) - f^*(w(h))$. Ou seja, mostra a diferença das densidades apresentadas na Figura 8 ao longo da escala do log dos rendimentos-hora. Comparando os dois gráficos podemos entender que há mais mulheres que homens com log-rendimentos negativos e uma maior concentração de homens nos log-rendimentos acima de 4.

A mudança na densidade é decomposta nos segundo e terceiro gráficos da Figura 9. O segundo gráfico ilustra a contribuição nas mudanças pelos atributos dos trabalhadores, ou covariáveis, para a mudança observada. Ele é obtido pela diferença entre a densidade marginal estimada das mulheres e a contrafactual com todos os atributos como dos homens: $f^*(w(m)) - f^*(w(m), x(h))$.

De forma similar, o terceiro gráfico, mostra a influência do retorno destes atributos para os trabalhadores, isto é os coeficientes, nas diferenças observadas das densidades dos gêneros. Isto é obtido pela diferença entre a densidade contrafactual com todos os atributos como dos homens e a densidade marginal estimada dos homens: $f^*(w(m), x(h)) - f^*(w(h))$. Segundo Guimarães (2001), neste último passo a contribuição dos coeficientes, não pode ser obtida utilizando a técnica da estimação semi-paramétrica desenvolvida por DFL. Em contrapartida, a metodologia de MM permite a identificação da parte que não é explicada pela mudança na distribuição das covariáveis, e pode ser atribuído à influência da mudança nos retornos ao capital humano e outras covariáveis, na mudança total.

Analisando a Figura 9 podemos verificar que os coeficientes explicam a maior proporção de mulheres em rendimentos mais baixos. Se eliminarmos o impacto dos coeficientes (último gráfico da figura), a diferença de densidades de rendimentos entre homens e mulheres seria como o gráfico do meio da Figura 9, ao invés do primeiro gráfico. Neste gráfico do meio, vemos que haveria uma menor concentração de mulheres nos log-rendimentos mais baixos se as mulheres tivessem os mesmos retornos aos atributos pessoais que os homens. Por outro lado, haveria mais mulheres nos rendimentos mais altos se os diferenciais de retornos fossem eliminados. A partir do log-rendimento 4 (ou aproximadamente R\$2.200,00 por mês) as covariáveis não têm influência na diferença de rendimentos. Já a diferença em favor das mulheres no intervalo do log dos rendimentos entre três e quatro (entre R\$800,00 e R\$2.200,00 por mês), por sua vez, é explicado pelos atributos dos trabalhadores (impacto covariáveis), mas também se verifica uma parte sendo devida ao retorno destes atributos. Deve-se destacar que o efeito negativo da discriminação não é monotônico, pois há o intervalo de rendimentos citado acima em que a eliminação dos diferentes retornos não irá aumentar a proporção de mulheres em relação aos homens.

Para completar a análise buscamos entender que atributos têm maior impacto nos diferenciais entre as densidades de rendimentos observados. Os gráficos seguintes da Figura 10 mostram os impactos dos retornos dos atributos dos trabalhadores separadamente. Estes gráficos são obtidos plotando a diferença entre a densidade contrafactual das mulheres com apenas o atributo correspondente distribuído segundo os homens e a densidade marginal estimada dos homens: $f^*(w(m), z(h)) - f^*(w(h))$. Sua interpretação é similar ao terceiro gráfico da Figura 9: valores positivos indicam que os retornos diferenciados para aquele atributo levam a um aumento na proporção de mulheres naquele nível de rendimentos. Valores positivos podem ser interpretados como uma discriminação na remuneração daquele atributo contra as mulheres.

O primeiro gráfico da Figura 10 representa o retorno do atributo idade. Os gráficos seguintes representam os retornos do atributo escolaridade subdividida por níveis, sendo que, a constante representa nenhum ou menos de um ano de escolaridade.

Podemos observar que a discriminação devida ao atributo idade é maior nos extremos da distribuição dos rendimentos.⁵ Por outro lado, há um tratamento melhor para as mulheres nos salários médios, embora este efeito seja relativamente pequeno.

O atributo escolaridade apresenta diferentes padrões de acordo com os níveis. No primeiro nível (constante) não podemos caracterizar a discriminação para um intervalo específico da distribuição. Há uma oscilação muito grande. Por outro lado, chamam a atenção como os retornos à escolaridade são diferenciados contra as mulheres para os salários mais baixos dentro dos grupos de 4-7 anos de escolaridade (ensino fundamental incompleto), 11-14 e mais de 15 anos de escolaridade (nível superior completo ou incompleto). Isto contribui para o deslocamento da distribuição das mulheres para a esquerda. Nestes níveis de escolaridade, uma redução no tratamento diferenciado entre homens e mulheres teria um impacto significativo para as mulheres de rendimentos mais baixos.

Outra característica a todos os níveis de escolaridade, em diferentes graus, é a presença de possível discriminação no final da distribuição. Isto é, o retorno do atributo escolaridade (para qualquer nível) apresenta discriminação em relação às mulheres nos rendimentos mais elevados, em particular para os níveis de médio incompleto e superior incompleto (8-10 anos e 11-14 anos, respectivamente). A diferença positiva das mulheres nos gráficos a partir do log-rendimento 3 pode caracterizar a presença de *glass ceiling*. Nestes níveis de rendimento, a equiparação dos retornos geraria uma redução da concentração de mulheres nestes níveis de renda, em relação aos homens e um provável aumento de seus rendimentos.

Em suma, vemos que os resultados obtidos indicam a existência de diferenciais de rendimentos que não podem ser explicados por diferenças de atributos observados, ao longo da distribuição de rendimentos de homens e mulheres. Estes diferenciais, que podem ser interpretados como limites superiores de discriminação contra as mulheres, não são monotônicos ao longo da distribuição, nem similares entre atributos. A heterogeneidade da discriminação entre atributos e em diferentes partes da distribuição de rendimentos não pode ser inferido a partir de diferenças de médias das distribuições de homens e mulheres trabalhadoras, ou seja, através da aplicação do método Oaxaca-Blinder de decomposição de diferenciais de rendimentos. Desta forma, por exemplo, podemos estudar a possibilidade de existência de limites ao crescimento de rendimentos femininos, pela comparação de distribuições de renda condicionais, sugerindo a existência do chamado *glass ceiling* contra as mulheres no Brasil. Em adição vemos que a redução da discriminação contra as mulheres pode ter o efeito de aumentar a desigualdade de rendimentos, pois os retornos aos atributos são mais heterogêneos entre os homens.

⁵ Deve-se enfatizar que os gráficos da figura 10 não relacionam o nível do atributo ao nível de rendimento, e sim o efeito do atributo na distribuição de rendimentos entre homens e mulheres. Por exemplo, uma diferença positiva no início do gráfico de idade não implica que os rendimentos das mulheres são mais baixos dado que elas são mais jovens do que os homens.

Comentários Finais

Este trabalho procurou analisar os conhecidos diferenciais de rendimentos por gênero no Brasil, explorando não o conhecido diferencial médio, mas sim a possibilidade de crescentes diferenciais ao longo da distribuição de rendimentos, o chamado *glass ceiling*. Para isso, utilizamos a metodologia de Machado e Mata (2005), que consiste na estimação da equação de rendimentos pelo método de regressão quantílica, que geram uma família de coeficientes para a construção das respectivas densidades de homens e mulheres. Baseado nas densidades estimadas conseguimos fazer uma decomposição das diferenças destas densidade em duas partes: uma relativas as características individuais dos trabalhadores e outra relativa aos retornos destas características. Esta metodologia permite entender melhor o diferencial de rendimentos médios atribuídos à discriminação na remuneração de atributos entre homens e mulheres no Brasil e lançar luz sobre o efeito da discriminação sobre a desigualdade de rendimentos entre trabalhadores no país de um novo ponto de vista.

Os dados no Brasil, para 1999, indicam que o diferencial de rendimentos-hora médio de 10% para os homens em relação às mulheres é bastante variável ao longo das distribuições de rendimentos. O diferencial nos quantis pode variar de menos de 5% para os níveis baixos de rendimentos e próximos ao percentil 80, para mais de 30% para os percentis mais altos. O diferencial é crescente até pouco depois da mediana, quando cai e volta a subir no final da distribuição. O crescimento no final da distribuição pode ser associado ao chamado *glass ceiling*. Nos Brasil, a questão interessante é o crescimento do diferencial até os rendimentos medianos e seu formato de U invertido, excluindo os rendimentos mais altos, onde o *glass ceiling* aparece com força.

Pelas estimativas de regressão quantílica, conseguimos separar o impacto nos rendimentos de homens e mulheres ao longo da distribuição de rendimentos para cada covariável, permitindo identificar qual atributo pessoal e em que parte da distribuição de rendimentos há as maiores diferenças entre homens e mulheres.

Para a idade encontramos retornos crescentes para as mulheres, enquanto que para os homens o comportamento é crescente até o sétimo decil, onde passa a decrescer. Mas apesar desta queda, a diferença do retorno da idade entre homens e mulheres é favorável aos homens com exceção dos extremos da distribuição.

Os retornos aos diferentes níveis de escolaridade apresentaram um padrão de comportamento no qual podemos enfatizar três pontos: a menor desigualdade de rendimentos, dentro dos níveis de escolaridades para mulheres, em relação aos homens; a similaridade de rendimentos para salários baixos entre homens e mulheres, dentro do nível de escolaridade, e o pequeno ganho de rendimentos para mulheres ao longo dos decis da distribuição feminina, quando comparado com a distribuição dos homens. Este rápido crescimento até próximo da mediana dos diferenciais dos retornos dos homens em relação às mulheres pode justificar o padrão de rendimentos identificado acima.

Existem diferenciais explicados apenas pelos coeficientes. E eles são bastante heterogêneos. A novidade é que estes diferenciais são positivas para as mulheres para rendimentos relativamente baixos, mas negativos para rendimentos mais altos.

De forma geral, encontramos retornos positivos para a educação na distribuição de rendimentos, mas conseguimos diferenciar estes impactos por níveis de escolaridade e as diferenças relativas ao gênero. Esta diferença de comportamento implica que a escolaridade das mulheres contribui para a menor dispersão da distribuição de rendimentos. Este resultado é consistente com o encontrado por outros autores que utilizaram a metodologia de Machado e Mata (2004) em outros países, como visto.

Outra consequência dos diferentes retornos por níveis educacionais, é que amostras com trabalhadores mais educados mostram maior dispersão do que amostras de indivíduos menos educados. Este último resultado nos proporciona algumas conclusões importantes porque: não necessariamente maior educação implica em mais igualdade na distribuição per capita da renda; e um ano adicional de escolaridade tem impactos diferentes sobre os rendimentos para homens e mulheres com o mesmo nível de educação.

Por fim, tivemos a avaliação da prevalência no mercado de trabalho brasileiro do fenômeno conhecido na literatura econômica como *glass ceiling*, isto é, a existência de uma barreira invisível para o crescimento do rendimento das mulheres nos níveis de rendimentos mais altos. No exercício contrafactual, ao gerarmos as densidades empíricas, constatamos uma queda abrupta na concentração de mulheres em relação aos homens no final da distribuição de rendimentos. Em contrapartida, quando geramos a densidade contrafactual das mulheres se elas tivessem a distribuição de rendimentos dos homens, essa queda da concentração para os níveis mais altos não se verifica.

Baseado nestes resultados obtidos construímos um gráfico de diferença das densidades empíricas, de forma que conseguimos visualizar claramente a mudança de comportamento da distribuição. A partir do decil 80 há um crescimento muito rápido da diferença relativamente aos outros quantis favorável aos homens.

A análise da discriminação propiciou averiguar como a discriminação (entendido como o diferencial não explicado de rendimentos) ocorre para cada característica dos trabalhadores. A hipótese inicial de discriminação contra a mulher se confirmou. A discriminação apresenta diferentes padrões para todos os níveis. Uma característica comum a todos os níveis de escolaridade é a presença de discriminação no final da distribuição, o que caracteriza a queda abrupta da concentração de mulheres na parte final da distribuição. Em síntese parece haver indícios suficientes de que ocorre *glass ceiling* no Brasil.

As estimativas de rendimentos construídas pela metodologia de Machado e Mata (2004), possibilitaram a construção de medidas de desigualdade, como o índice L de Theil. As mulheres apresentaram distribuição mais equitativa que os homens. Os resultados obtidos mostraram maior desigualdade de rendimentos para homens que mulheres. Mas o exercício mais importante foi obtido pelo índice de Theil calculado através do exercício contrafactual para o rendimento das mulheres. Com base nos rendimentos das mulheres caso estas tivessem a distribuição dos homens, encontramos um novo índice de Theil maior que os obtidos anteriormente, o que também gerou um aumento no Theil entre-grupos. Ou seja, o desejo da redução da desigualdade entre gêneros pode gerar um conflito pelo desejo da redução da desigualdade da economia como um todo.

Como sugestão de pesquisas futuras, está uma análise temporal dos diferenciais de rendimentos ao longo de toda a distribuição. Por exemplo, pode-se identificar em que maneira a maior participação feminina ao longo dos anos 80 e 90 e a redução do diferencial médio de rendimentos afetaram a distribuição de rendimentos femininos. O aumento de rendimentos médios femininos pode ter sido devido ao aumento de toda a distribuição ou devido ao aumento dos rendimentos femininos mais altos.

REFERÊNCIAS BIBLIOGRÁFICAS

- AIGNER, D. J. & CAIN, G. G. (1977). **Statistical theories of discrimination in labor market**. *Industrial and Labor Relations Review*, 30:175–187.
- ALBRECHT, J., BJÖRKLUND, A. and VROMANIS, S. (2001). **There a Glass Ceiling in Sweden?** Mimeo.
- BARROS, R. P., RAMOS, L. (1992). **Gender: Differences in brazilian labor markets**. *Anais do XX Encontro Nacional de Economia*: 63–80.
- BARROS, R. P. e MENDONÇA, R. S. P. (1995). **Os determinantes da desigualdade no Brasil**. Rio de Janeiro: Texto para Discussão, n. 377.
- BECKER, G. (1957). **The Economics of Discrimination**. The University of Chicago Press.
- BORJAS, George. (1996). **Labor Economics**. New York, MacGraw-Hill.
- BULOW, Jeremy and SUMMERS, Lawrence (1986). **A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination and Keynesian Unemployment**. *Journal of Labor Economics*, 4, pp. 376-414.
- BUCHINSKY, M. 1998. **Recent advances in quantile regression models - A practical guideline for empirical research**. *Journal of Human Resources* 33 (1): 88-126.
- CAIN, Glen C. (1976). **The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey**. *Journal of Economic Literature*, 14, p. 1215-1257.
- CAVALCANTI T. e GUIMARÃES J. F. (1999). **Labor Market Discrimination: Quantile Regression Approach**. *Anais XXI Encontro Brasileiro de Econometria*, pp. 198-218.
- DINARDO J., FORTIN N. M. AND LEMIEUX T. (1996). **Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach**. *Econometrica*. 64(5). September, pp. 1001-44.
- delaRICA, DOLADO, J. and LLORENS, V. (2005). **Ceilings and Floors: Gender Wage Gaps by Education in Spain**. IZA Discussion paper 1483.
- EHREMBERG, R.G. e SMITH, R.S. (2000). **Moderna Economia do Trabalho: Teoria e Política Pública**. Rio de Janeiro, Makron.
- FITZENBERGER, B. and KURZ C. (1997). **New Insights on Earnings Trends Across Skill Groups and Industries in West Germany**, Mimeo, Universitat Konstanz.
- GUIMARÃES J. F. (2001). **Changes in the earning Distribution in Brazil, 1986-1995: Human Capital and Institutional Factors**. *Anais XXIII Encontro Brasileiro de Econometria*.
- GIUBERTI, Ana Carolina e MENEZES-FILHO, Naércio (2005). **Discriminação de rendimentos por gênero: uma comparação entre o Brasil e os Estados Unidos**. *Economia Aplicada*, vol.9, no.3, p.369-384.
- GREENE, Willian H (2000). **Econometric analysis**. Fourth Edition, New York

University.

KASSOUF, A. L. & SILVA, N. (2000). **Mercados de trabalho formal e informal: Uma análise da discriminação e da segmentação**. Nova Economia Aplicada, 10(1):41–77. Revista do Departamento de Ciências Econômicas da UFMG.

KOENKER R and BASSET G (1978). **Regression quantiles**. *Econometrica* 46:33–50.

KOENKER R and BASSET G (1982). **Robust tests for heteroscedasticity based on regression quantiles**. *Econometrica*, 50, 43-61

KOENKER, Roger (2005). **Quantile Regression**. Cambridge: CUP.

LOUREIRO, Paulo R. A. (2003) **Uma resenha teórica e empírica sobre economia da discriminação**. Revista Brasileira de Economia, Mar 2003, vol.57, no.1, p.125-157.

MACHADO J.A.F., MATA J (2005). **Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression**. *Journal of Applied Econometrics*.

OAXACA, Ronald. (1973). **Male-Female Wage Differentials in Urban Labor Markets**. *International Economic Review*, 14 (3): 693-709.

OMETTO, A. M., HOFFMAN, R., & ALVES, M. C. (1999). **Participação da mulher no mercado de trabalho: Discriminação em Pernambuco e São Paulo**. Revista Brasileira de Economia, 53(3):287–322.

PIORE, Michael (1970). **Jobs and Training: Manpower Policy, The State and the Poor**, ed. S. Beer and R. Barring. Winthrop Press.

PHELPS, E. S. (1972). **The statistical theory of racism and sexism**. *American Economic Review*, LXII: 659–661.

RAMOS, Lauro, VIEIRA, Maria Lucia (1996). A Relação entre Educação e Salários no Brasil. In: A Economia Brasileira em Retrospectiva - 1996. **Rio de Janeiro: Ipea**.

SPENCE, M. (1973). **Job market signaling**. *Quarterly Journal of Economics*, LXXXVII: 355–74.

Figura 1 – Remuneração Média a preços de Dez/03, segundo Gênero e Grau de instrução Brasil - 2002

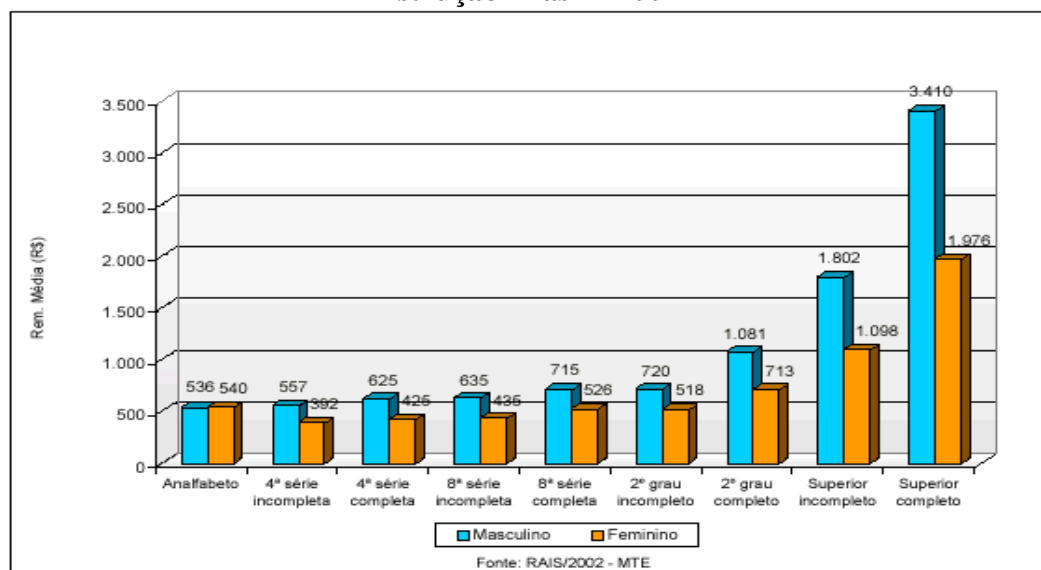


Figura 2 – Diferença de rendimentos entre homens e mulheres

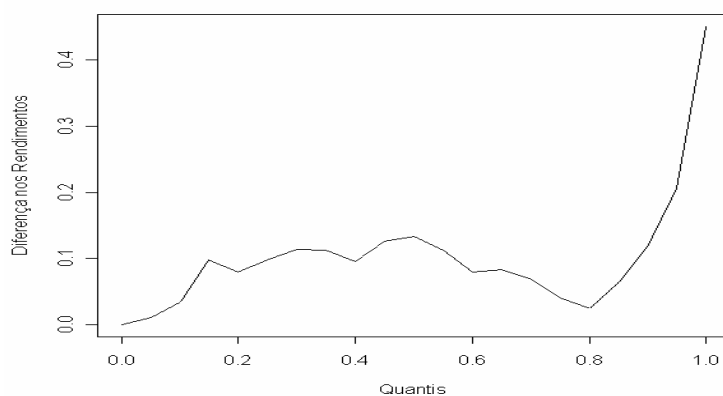


Figura 3 – Box Plot Rendimento dos Gêneros

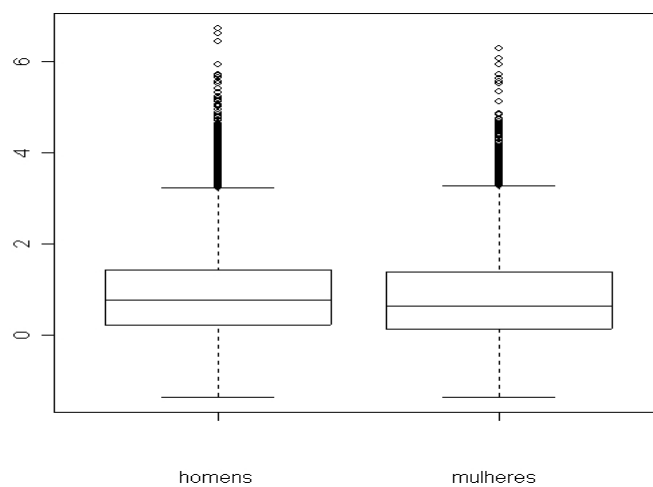


Figura 4 - Regressão Quantílica para uma distribuição com erros heterocedásticos

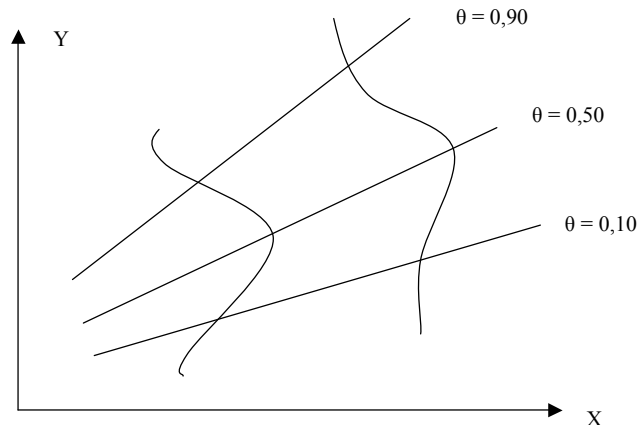


Figura 5 - Estimativas da Regressão Quantílica (intercepto e idade)

Os gráficos abaixo apresentam as estimativas para cada covariável individual indicada. As linhas pontilhadas representam o intervalo de confiança de 95% e a linha horizontal a estimativa por mínimos quadrados ordinários. Os gráficos à esquerda mostram as estimativas para as mulheres, dos homens no centro e a diferença entre homens e mulheres à direita.

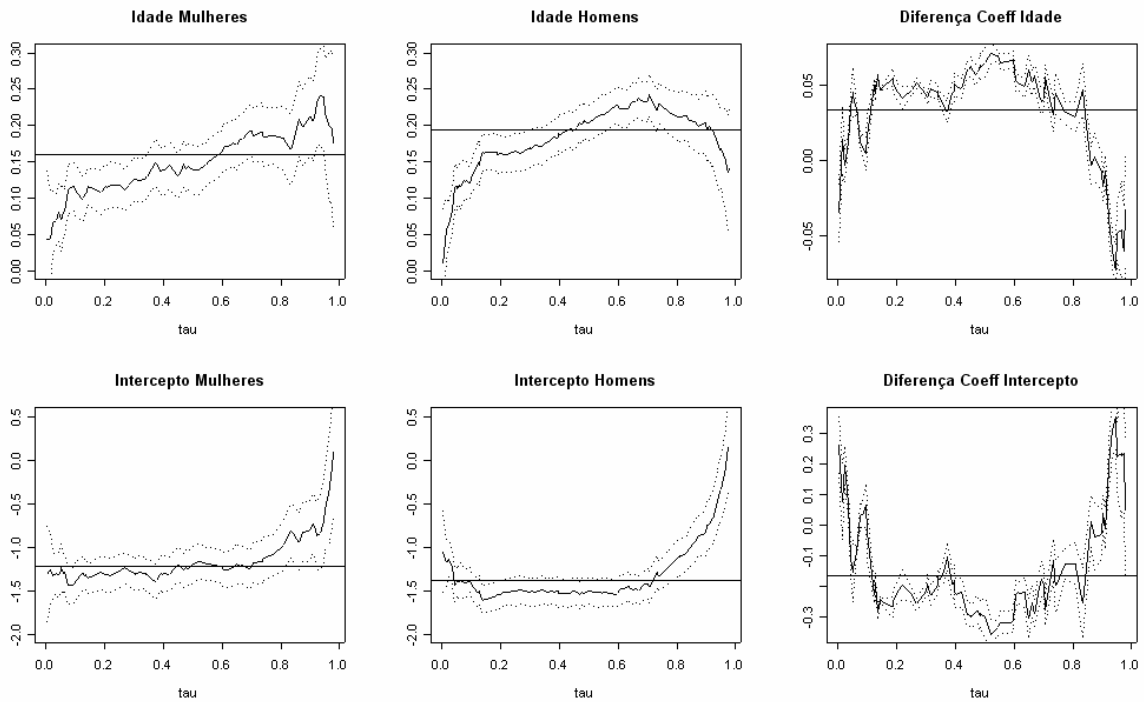


Figura 6 - Estimativas da Regressão Quantílica (níveis educacionais)

Os gráficos acima apresentam as estimativas para cada covariável individual indicada. As linhas pontilhadas representam o intervalo de confiança de 95% e a linha horizontal a estimativa por mínimos quadrados ordinários. Os gráficos à esquerda mostram as estimativas para as mulheres, dos homens no centro e a diferença entre homens e mulheres à direita.

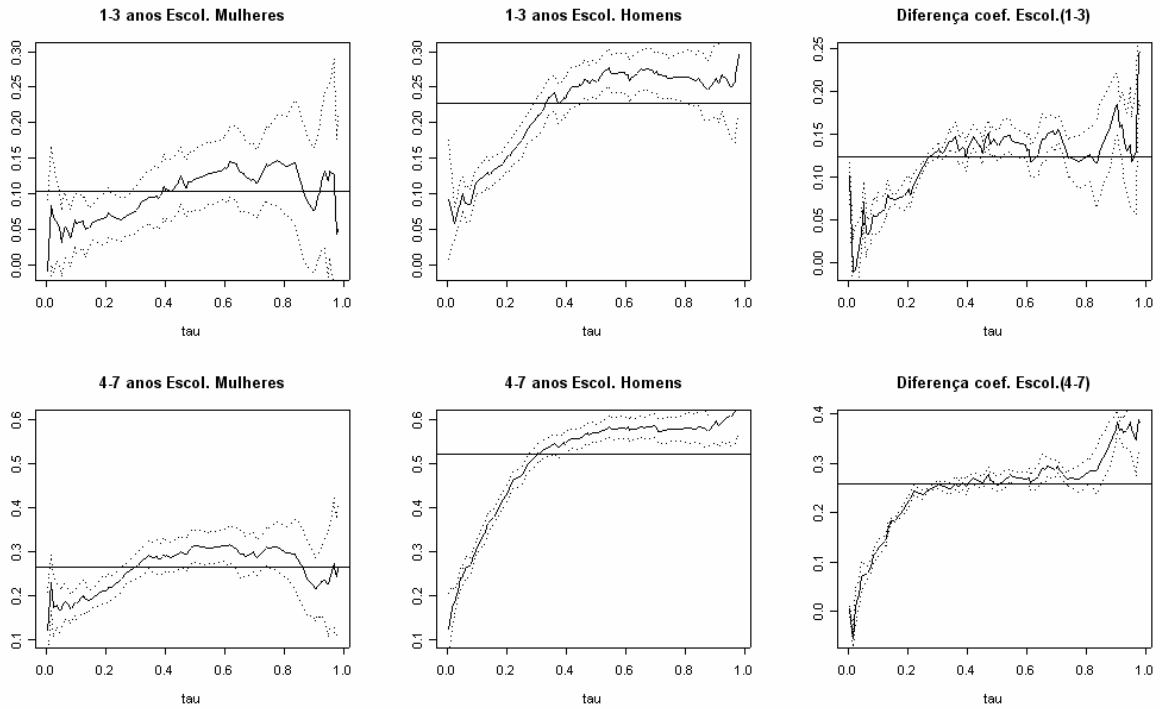


Figura 7- Estimativas da Regressão Quantílica (níveis educacionais)

Os gráficos acima apresentam as estimativas para cada covariável individual indicada. As linhas pontilhadas representam o intervalo de confiança de 95% e a linha horizontal a estimativa por mínimos quadrados ordinários. Os gráficos à esquerda mostram as estimativas para as mulheres, dos homens no centro e a diferença entre homens e mulheres à direita.

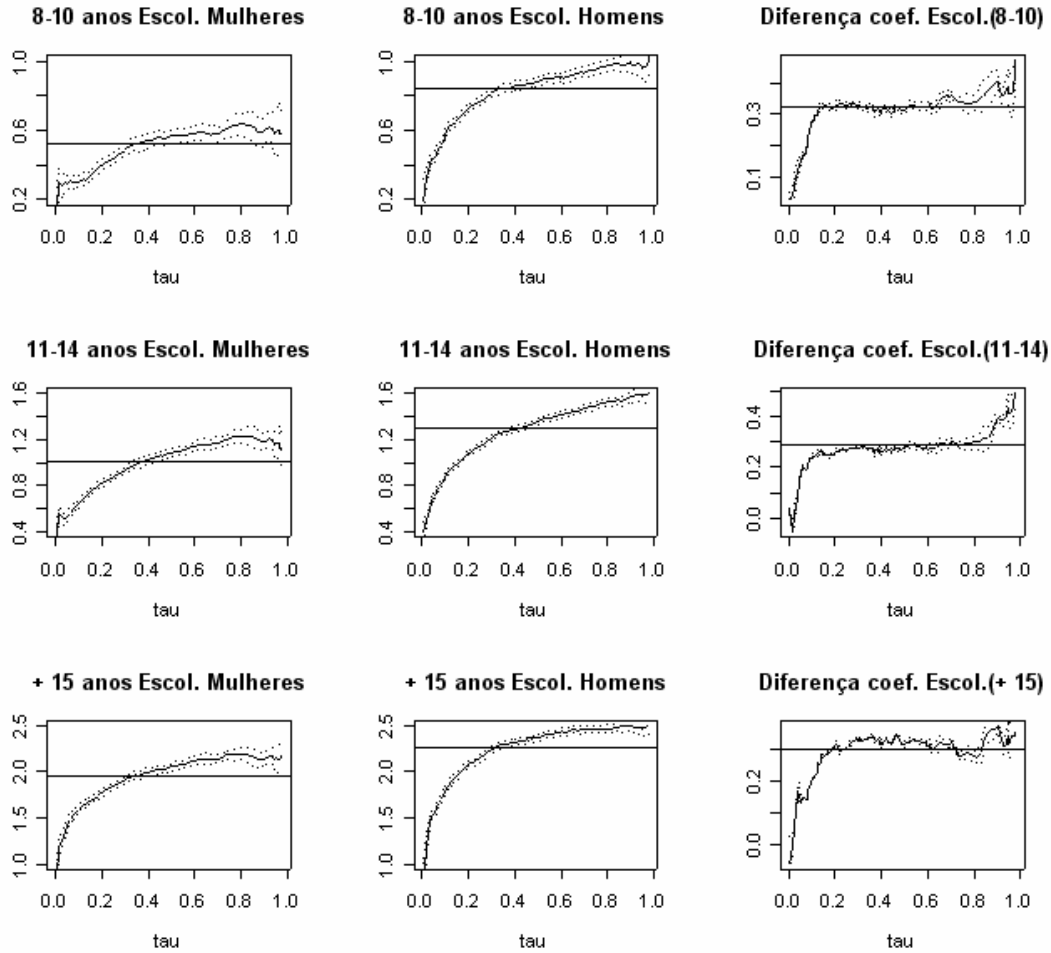


Figura 8 - Densidades Empíricas

Densidades empíricas dos homens (linha cheia), mulheres (linha pontilhada) e mulheres com a função de distribuição dos rendimentos dos homens (linha pontilhada mais fina). Log dos rendimentos-hora no eixo horizontal

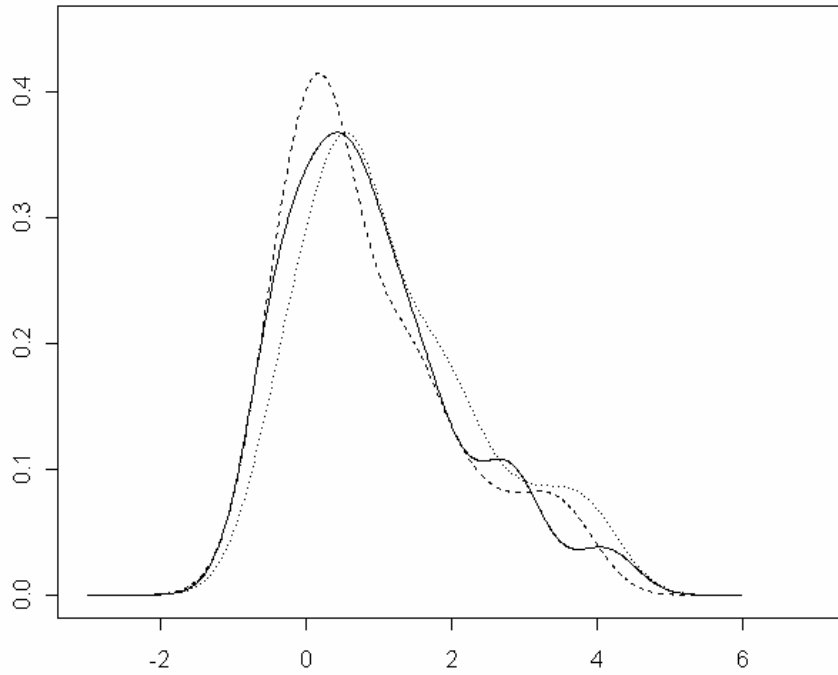


Figura 9 - Mudanças nas Densidades

A base de comparação se refere à diferença entre mulheres e homens ($f^*(w(m)) - f^*(w(h))$)

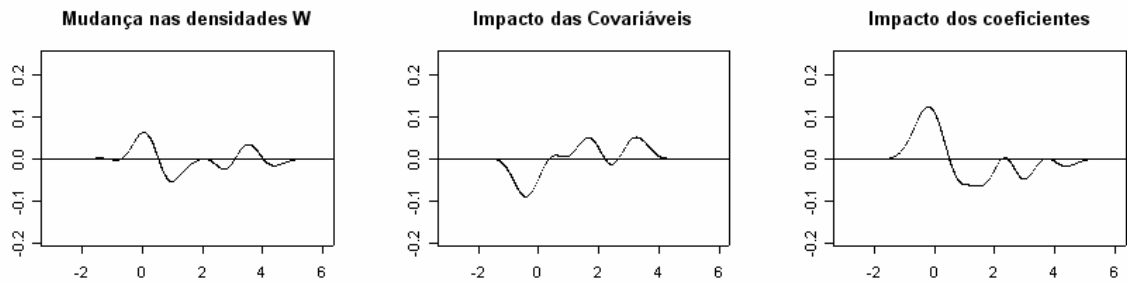
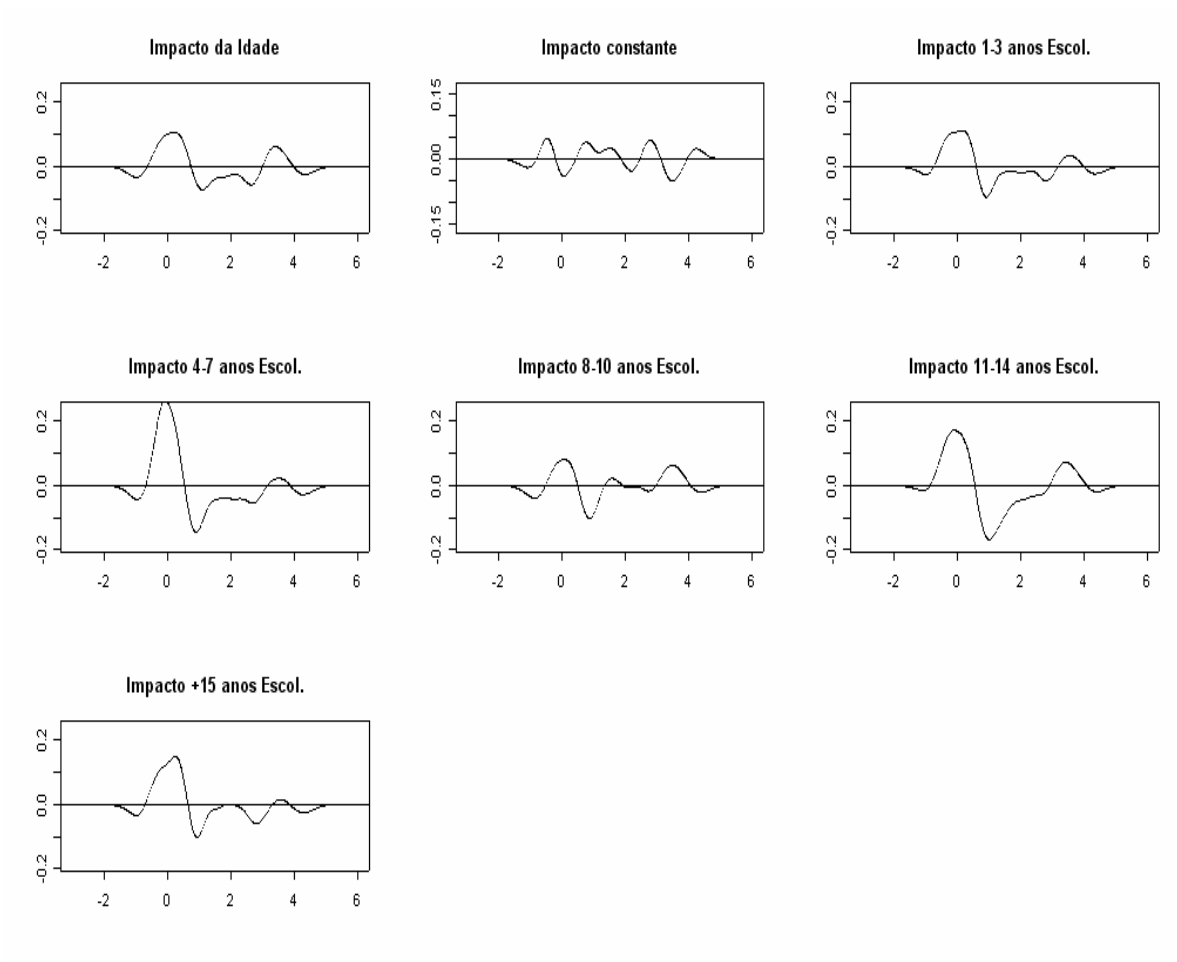


Figura 10 – Retorno dos atributos individuais



Anexo: Detalhamento do método de Machado e Mata (2005)

A.1. Distribuição Condicional dos Rendimentos

Nosso objetivo é entender o papel das características pessoais nos diferenciais de rendimentos entre homens e mulheres. Inicialmente estimamos a distribuição condicional dos rendimentos segundo as covariáveis de interesse, o qual obtemos usando a técnica de regressão quantílica desenvolvida por Koenker e Basset (1978). O primeiro passo é especificar o τ -ésimo quantil condicional da distribuição do (log) rendimento por hora w em um vetor X dos atributos dos trabalhadores como $Q_w(\tau|x)$ para $\tau \in (0, 1)$. O conjunto de atributos individuais X usado na análise empírica consiste na informação sobre a idade e escolaridade. Baseado nisto nos especificamos a função condicional quantílica como: $Q_w(\tau|x) = x'\beta(\tau)$, onde $\beta(\tau)$ é o coeficiente da regressão quantílica a ser estimado.

A proposta de MM consiste num método de decomposição da mudança da distribuição de rendimentos entre os gêneros em vários fatores que contribuem para essas mudanças. O método é baseado na estimação das distribuições marginais dos rendimentos consistentes com uma distribuição condicional estimada pela regressão quantílica assim como com alguma distribuição hipotética para as covariáveis.

A decomposição contrafactual das mudanças nas densidades dos rendimentos é obtida por médias das regressões quantílicas e pelo método de reamostragem, segundo Machado e Mata (2004), ela pode ser dividida em quatro passos: (1) a estimação da distribuição dos rendimentos condicional; (2) a estimação das densidades marginais resultantes do modelo pra ambos gêneros; (3) as densidades contrafactuais; (4) a decomposição das mudanças segundo a contribuição dos coeficientes, covariáveis e um resíduo. Os tópicos seguintes explicam detalhadamente cada passo.

A2. Densidade Marginal dos Rendimentos

Esta parte da estimação é projetada para obter a densidade marginal dos rendimentos baseado na distribuição condicional estimada anteriormente. Para melhor compreender este processo observe que segundo MM, a densidade marginal da variável dependente poderia ser facilmente obtida dos dados, mas não seria condicionada nas covariáveis especificadas acima e conseqüentemente não permitiria que nós prosseguíssemos com o exercício contrafactual.

O ponto chave do método usa o Teorema da Probabilidade Inversa. Seja $U \in (0, 1)$ uma variável aleatória uniforme, então $F^{-1}(U)$ tem distribuição F . Assim dado uma amostra de $\tau_1, \tau_2, \dots, \tau_n$ de uma distribuição $U(0, 1)$ tal que os quantis condicionais dos rendimentos dado x , $\{x'\beta(\tau)\}_{i=1}^n$ consiste de uma amostra aleatória da distribuição condicional dos rendimentos dado $X = x$, estimado no primeiro passo. Aqui projetamos a escolha arbitrária de um X específico pelo método de amostragem com reposição de uma amostra de covariáveis de uma distribuição apropriada. O algoritmo pode ser descrito da seguinte forma:

Considere os dois gêneros, homens (h) e mulheres (m), tal que $G = m, h$ e $w(G)$ e

$x(G)$ denotam os rendimentos e as p variáveis para cada gênero, respectivamente. Além disso, $g(x; G)$ é a distribuição conjunta das variáveis no gênero G . O algoritmo então consiste em:

1. Geramos uma amostra aleatória de tamanho n de uma $U[0, 1]$: $\tau_1, \tau_2, \dots, \tau_n$.
2. Para cada τ_j estimamos $Q_{wh}(\tau_j/X)$ e $Q_{wm}(\tau_j/X)$ para $j = 1, \dots, n$ produzindo n estimativas dos coeficientes de regressão nos quantis $\hat{\beta}_h(\tau_j)$ e $\hat{\beta}_m(\tau_j)$.
3. Geramos uma amostra com reposição de tamanho n das linhas de X_h e X_m e denotamos por $\{X_h^*(j)\}_{j=1}^n$ e $\{X_m^*(j)\}_{j=1}^n$.
4. Computamos $\{w_h^*(j) \equiv X_h^*(j)' \hat{\beta}_h(\tau_j)\}_{j=1}^n$ e $\{w_m^*(j) \equiv X_m^*(j)' \hat{\beta}_m(\tau_j)\}_{j=1}^n$.

A idéia básica é de retirar uma amostra aleatória de variáveis explicativas X usando amostragem com reposição. Conseqüentemente, podemos gerar uma amostra aleatória de tamanho n de nosso coeficiente de discriminação de uma distribuição apropriada, onde w^* é uma amostra aleatória da distribuição de rendimentos, construída a partir dos estimadores de regressão quantílica e da amostra aleatório de X .

A.3. Densidade Contrafactual

O primeiro exercício contrafactual corresponde à estimação da função de densidade das mulheres se todas as covariáveis fossem distribuídas como dos homens, e os trabalhadores fossem pagos de acordo com a estrutura dos rendimentos das mulheres. Para isso, seguimos o algoritmo descrito acima até o terceiro item. No quarto item, ao invés de extrairmos a amostra das linhas da matriz de covariáveis das mulheres, $X(m)$, agora fazemos uma amostragem com reposição das linhas da matriz de covariáveis dos homens, $X(h)$.

O segundo tipo de exercício consiste em estimar a densidade das mulheres se apenas uma das covariáveis fosse distribuída como dos homens. Este exercício não é obtido diretamente como o primeiro. Considere uma covariável específica qualquer, denotemos $z(G)$, por exemplo, um nível de escolaridade qualquer. Nos queremos estimar a distribuição dos rendimentos das mulheres caso uma covariável $z(G)$ fosse distribuída como a distribuição dos homens e as outras demais covariáveis fossem distribuídas como das mulheres. O algoritmo para este segundo tipo de estimação contrafactual consiste em:

1. Seguimos os itens de 1 a 4 do algoritmo anterior para gerar uma amostra de tamanho n das densidades dos rendimentos para as mulheres: $\{w_i^*(m)\}_{i=1}^n$.
2. Pegue uma classe $C_1(m)$ (digamos escolaridade)
 - (a) Selecionamos um subconjunto da amostra aleatória dos rendimentos gerado no passo 1 correspondente ao espaço $I_1 = \{i = 1, \dots, n \mid z_i(m) \in C_1(m)\}$.

$C_i(m)$ isto é $\{w_i^*(m)\}_{i \in I_i}$

(b) Geramos uma amostra aleatória de tamanho $n \times f_j(m)$ com reposição de $\{w_i^*(m)\}_{i \in I_i}$.

3. Repetimos o passo 2 para $j = 2, \dots, j$.

4. Montamos todas as amostras dos rendimentos juntas em relação às classes.

Formalmente a densidade dos rendimentos para as mulheres caso um nível específico de escolaridade tivesse sido distribuído como nos homens seria dado por: $\int f(w | z, G_w = m) dF(z | G_z = h)$ onde G representa o gênero e o subscrito a variável de interesse. Assim, $f(w/z, G_w = m)$ denota a distribuição dos rendimentos das mulheres para a covariável z , e $F(z | G_z = h)$ é a distribuição da covariável z para os homens. A distribuição marginal dos rendimentos das mulheres pode, portanto ser escrita como uma composição da distribuição dos rendimentos pra trabalhadores com e sem este nível de escolaridade (z) com pesos iguais a proporção de homens trabalhadores com e sem este nível de escolaridade. Como em DiNardo, Fortin e Lemieux (1996), nos podemos manipular estes pesos, baseado na sua frequência, a fim de estimar a diferença das densidades contrafactuais.