



1290003778

TCC/UNICAMP  
G745i  
1290003778/IE

ESTHER MAJEROWICZ GOUVELA

**O IMPACTO DA SINDICALIZAÇÃO NA RENDA  
DO TRABALHO: UMA APLICAÇÃO DA TÉCNICA DE  
*PROPENSITY SCORE***

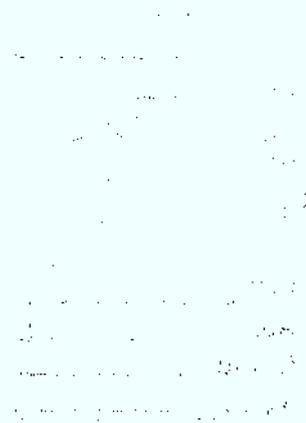
CAMPINAS  
2008

TCC/UNICAMP  
G745i  
1290003778/IE

CEDOC/IE/UNICAMP

829006208

ESTHER MAJEROWICZ GOUVEIA



# **O IMPACTO DA SINDICALIZAÇÃO NA RENDA DO TRABALHO: UMA APLICAÇÃO DA TÉCNICA DE *PROPENSITY SCORE***

Monografia apresentada ao Instituto de  
Economia, da Universidade Estadual de  
Campinas, para conclusão do curso de  
Bacharelado em Ciências Econômicas.  
Área de concentração: Econometria.

Orientador: Prof. Dr. Alexandre Gori  
Maia.

## Resumo

Uma vez que grande parte dos estudos sobre os quais os economistas se debruçam possui caráter observacional, os pesquisadores da área recorrentemente encontram-se limitados no que diz respeito ao instrumental analítico para a produção de dados empíricos. Para melhor compreender as implicações de tais limitações e tentar contorná-las, o presente trabalho propõe-se a estudar o impacto da sindicalização na renda do trabalhador através da aplicação de uma metodologia de pareamento, o propensity score. A pesquisa baseou-se na revisão bibliográfica sobre o histórico do movimento sindical no Brasil e sobre a avaliação de tratamentos e na análise estatística. A construção do panorama geral da história do movimento sindical no Brasil permitiu explicitar que a relação existente entre os sindicatos e a renda do trabalho não se dá de forma constante e que a última pode ser uma boa proxy da luta sindical. A exposição dos desenhos experimentais evidenciou formas de controlar possíveis fontes de vieses estatísticos e apresentou o propensity score, uma metodologia de pareamento em estudos observacionais, como uma solução metodológica possível para o caso da sindicalização na renda do trabalho. Finalmente, o trabalho, ao aplicar tal metodologia, pôde demonstrar seu funcionamento e comparar como as suas estimativas diferem daquelas produzidas pelo MQO e da análise das medidas-resumo.

**Palavras-chaves:** *propensity score*; distribuição rendimentos; pareamento dados; sindicato;

## Sumário

<b>Introdução.....</b>	<b>5</b>
<b>Capítulo 1: Breve histórico da organização sindical no Brasil.....</b>	<b>7</b>
<b>Capítulo 2: Avaliação de tratamentos .....</b>	<b>10</b>
2.1 Planejamento de experimentos.....	12
2.2 Uma proposta de desenho <i>quasi</i> -experimental.....	15
2.3 O pareamento e o <i>propensity score</i> .....	17
<b>Capítulo 3: O caso da sindicalização sobre o rendimento do trabalho.....</b>	<b>24</b>
<b>Conclusão .....</b>	<b>41</b>
<b>Referências .....</b>	<b>43</b>

## Introdução

Pensar na economia de uma região ou nação implica buscar entender como os seus agentes sociais atuam na realidade para mudar ou perpetuar o curso do arranjo econômico estabelecido. Tais agentes, entendidos como o Estado, as classes e as frações de classe, estão situados em cenários com disposição ímpar dos mercados, das políticas e da legislação, e baseiam-se em diferentes pressupostos teóricos para intervir na sociedade em busca de atingir os resultados e objetivos que desejam. Desta forma, avaliar a coerência, a adequação e o papel de suas ações em relação ao que julgam estar promovendo é, em parte, identificar e medir o impacto das mesmas. De igual maneira, pode-se avaliar o resultado da existência de determinada característica do cenário econômico sobre algum aspecto da sociedade para melhor entendê-la.

Entretanto, as ciências econômicas, assim como as ciências humanas em geral, recorrentemente se enfrentam com circunstâncias não tão comuns as demais ciências no que tange a tarefa da mensuração de impactos. Isto ocorre porque, em geral, os economistas se deparam com dados observacionais onde as ferramentas ideais de mensuração não podem ser utilizadas visto que seus pressupostos não são verificados.

Uma vez que a produção de dados empíricos sobre as questões econômicas é de extremo valor para poder forjar, legitimar ou refutar as teorias econômicas é que cabe discutir a qualidade de tais dados em termos de aproximação da realidade. De tal maneira, torna-se necessário ao saber econômico compreender os limites e a adequação das técnicas de produção de dados empíricos de forma que ambos os lados de uma controvérsia estejam qualificados para defender as suas posições e que a última, se não puder ser resolvida, ao menos seja poupada de gravitar em torno de aspectos errôneos ou secundários.

Assim sendo, para poder exemplificar a natureza recorrente dos problemas com que as ciências econômicas se deparam no desafio de mensuração de impactos é que o presente trabalho se propõe a estimar o efeito da sindicalização na renda do trabalho no Brasil para o ano de 2006. Para tal feito, o trabalho foi estruturado em três capítulos. No primeiro deles pretendeu-se descrever sucintamente os movimentos gerais da história do sindicalismo brasileiro e evidenciar seus principais marcos para assim situar ao leitor o ponto em que se insere a análise estática dos dados. O segundo capítulo trata das questões relativas ao planejamento de experimentos visto que para a produção de bons dados é fundamental a utilização de um desenho experimental

robusto e que seja adequado ao fenômeno em estudo. Nesta parte do trabalho propõe-se o *propensity score* como uma das soluções possíveis para avaliar o impacto da sindicalização na renda do trabalho. Por fim, o último capítulo demonstra como foi aplicada tal técnica aos dados da Pesquisa Nacional por Amostra de Domicílios 2006 (PNAD) e aponta os dados obtidos por tal procedimento.

## **Capítulo 1: Breve histórico da organização sindical no Brasil**

Existem muitas formas de organização da classe trabalhadora e o sindicato é uma concretização específica de tais formas e sua importância histórica varia de nação para nação. Os sindicatos têm seu berço na Inglaterra com o nome de *Trade Unions* ou *Labour Unions*. Eles surgem a partir da necessidade dos trabalhadores de se contraporem as péssimas condições de vida impostas pela nascente sociedade capitalista trazida pela Revolução Industrial. Por muito tempo irão atuar clandestinamente, utilizando a greve como principal instrumento na luta pela redução da jornada de trabalho, por aumentos salariais e por melhores condições de trabalho.

Desde o surgimento das *Trade Unions* na Inglaterra até os dias de hoje vem sendo debatido qual seria o seu potencial de transformação na sociedade. Para muitos a luta sindical se limita ao âmbito das reivindicações economicistas, servindo somente como um dique ao aumento da exploração capitalista ou a melhoras nesta condição. Para outros, o sindicato pode servir como ferramenta da classe trabalhadora, pólo do poder dual, para a superação do capitalismo e para a construção da sociedade socialista.

No Brasil, a classe operária começa a se desenvolver no final do século XIX como resultado de inúmeras transformações que ocorriam no país. A economia do café enquanto modelo agrário-exportador ganha força ao se deslocar para o Oeste Paulista criando as condições para o estabelecimento do capital industrial e para o trabalho assalariado. Desta forma, a mão-de-obra européia toma lugar da escrava e os primeiros núcleos operários surgem principalmente em São Paulo e no Rio de Janeiro. Este nascente operariado é submetido a difíceis condições de vida e trabalho tendo que assumir uma jornada de trabalho que variava de 12 a 15 horas por dia sem nenhum dia de repouso e desamparado de qualquer forma de contrato legal. É a partir desta situação que surgem as primeiras formas de organização dos trabalhadores que tinham como finalidade dar ajuda recíproca e promover obras assistenciais, eram as chamadas sociedades de socorro e ajuda mútua. Destas organizações surgem as sociedades de resistência que dão origem aos sindicatos. O movimento operário deste período é denominado por muitos autores como anarco-sindicalismo, reflexo dos ideais trazidos pelos imigrantes europeus que já possuíam experiência organizativa em seus países de origem.

Entretanto, ocorrem mudanças significativas na organização sindical que passa a ter caráter paraestatal a partir da Revolução de 1930. Em 1931 o recém criado Ministério do Trabalho promulga o decreto de sindicalização de patrões e empregados. A legislação social do

período possui quatro núcleos básicos: a legislação previdenciária, as leis trabalhistas, a legislação sindical que institui a unicidade sindical e legitima a intervenção estatal no seio dos sindicatos e a instituição da Justiça do Trabalho. Em 1940 é criado o imposto sindical e três anos depois é promulgada a CLT (consolidação da legislação trabalhista). Tendo em vista as transformações vividas no período e a proibição de greves, o movimento sindical passa a atuar pelos canais institucionais e só em 1955 é que nota-se uma nova expansão culminando em 1961 com a criação do Comando Geral dos Trabalhadores.

Com o golpe militar de 1964 houve uma enorme repressão da luta e da organização sindical e 1966 o Fundo de Garantia do Tempo de Serviço (FGTS) é instituído para substituir a estabilidade empregatícia. É somente a partir de meados da década de 70 que o movimento sindical volta ao cenário como importante ator na luta dos trabalhadores e pela redemocratização do país. Este período é o que prepara o movimento sindical para o que, do ponto de vista da mobilização e organização, é considerado como o seu “momento de ouro” na primeira metade da década de 80.

Em 1978, o sindicato dos metalúrgicos do ABC paulista sob a presidência de Luís Inácio Lula da Silva entra em greve e é seguido por outras categorias resultando em um período de ascenso das lutas e da capacidade de reivindicação dos trabalhadores. Como desdobramento deste processo os trabalhadores fundam o Partido dos Trabalhadores (PT) e em 1983 a Central Única dos Trabalhadores (CUT). Posteriormente são criadas a Confederação Geral dos Trabalhadores (CGT) em 1986 e a Força Sindical em 1991 que não se alinham politicamente com o sindicalismo promovido pela CUT.

Entretanto, com o advento da reestruturação produtiva, das reformas neoliberais, da globalização e com a queda do muro de Berlim, as organizações sindicais sofrem duros golpes e o eixo do debate sobre o sindicato se altera: abandona-se a perspectiva classista e inicia-se o questionamento da própria validade da luta sindical. Além da busca constante da individualização das relações de trabalho pela política neoliberal, este questionamento também tem como uma de suas raízes o enorme aumento do mercado de trabalho informal. Deste modo, os sindicatos passam a expressar frações decrescentes da classe trabalhadora na mesma proporção em que se esvazia o mercado formal. Do ponto de vista teórico surgem formulações sobre o fim da história e o fim das classes sociais.

Com todas essas transformações econômicas, políticas e ideológicas torna-se importante dar suporte empírico para melhor compreender qual é o papel do sindicato no Brasil atualmente. Uma boa proxy para captar o seu poder de intervenção na sociedade é a renda do trabalho, um dos aspectos fundamentais da condição de vida do trabalhador brasileiro. Assim, medir o impacto da sindicalização na renda do trabalho é uma forma de dar mais elementos para o debate e assim qualificar a discussão em torno da legitimidade da luta sindical, seus possíveis horizontes e também de suas perspectivas de transformação social.

## Capítulo 2: Avaliação de tratamentos

Para mensurar o impacto da sindicalização sobre a renda do trabalho, é essencial compreender o planejamento de um estudo que visa avaliar o efeito de tratamentos sobre alguma variável de interesse. Por tratamento entende-se qualquer ação ou elemento da realidade sobre o qual é exposta uma amostra de alguma população. Assim, avaliar o tratamento consiste em medir os efeitos de tal exposição sobre algum aspecto de interesse que o pesquisador tenha na amostra. Este termo provém da área médica onde, em geral, as ações que se pretendem discutir se apresentam na forma de regime de tratamento. Entretanto, em economia o termo intervenção seria igualmente adequado.

Existem várias maneiras de planejar um experimento para medir o impacto de um tratamento e o desenho experimental ideal varia de acordo com o problema que está sendo proposto. O sucesso de um pesquisador na tentativa de avaliar o impacto de uma intervenção está em muito relacionado com a capacidade do desenho experimental adotado de controlar os outros fatores que podem afetar a variável (Y) sobre a qual se pretende medir o impacto. Este aspecto é importante, pois se os outros fatores não forem controlados seus efeitos podem ser confundidos com os da intervenção (X) e levar o pesquisador a falsas conclusões sobre os impactos de X. Tal preocupação diz respeito à validade interna do experimento a qual, por sua vez, refere-se à coerência do experimento em avaliar o impacto de X para a amostra em estudo, ou seja, em sua capacidade de isolar os efeitos de X e medi-los adequadamente no grupo que sofreu a exposição.

Por outro lado, construir um experimento bem sucedido significa poder tirar conclusões a partir da amostra que possam ser aplicadas ao conjunto da população. Tendo em vista que na maioria dos casos o que de fato interessa é afirmar algo sobre determinada população, o experimento com uma amostra de pouco serve se não puder ter seus resultados generalizados. Ainda que possua validade interna e faça afirmativas válidas para a amostra é preciso que tenha validade externa e possa oferecer algum conhecimento sobre a população.

Desta forma, é importante conhecer os principais fatores que podem ameaçar a validade interna e externa de um experimento e como planejar um desenho experimental adequado para controlar tais ameaças. Do ponto de vista da validade interna, CAMPBELL (1966) elenca oito principais ameaças, a saber:

- i) história: entre o momento inicial e o momento final do experimento, em que são feitas as mensurações, eventos históricos que alterem o objeto de estudo podem ocorrer e seus efeitos serão somados ao efeito de X;
- ii) maturação: são os efeitos produzidos pela passagem do tempo em si. Em geral, estes estão associados a fenômenos biológicos ou psicológicos (aumento da fadiga, envelhecimento, etc.) e se constituem como hipótese rival a X na explicação das transformações ocorridas na amostra;
- iii) instrumentação: se refere as interferências que podem ser produzidas pelos instrumentos de mensuração, como uma mudança na calibragem ou o uso de observadores diferentes. Ao utilizar padrões de mensuração distintos, diferenças seriam produzidas que poderiam ser atribuídas enganosamente a X;
- iv) testagem: muitos experimentos se valem de um processo de pré-testagem da amostra e este processo por si mesmo pode mudar o objeto que está sendo testado. Quando se trata da testagem com pessoas, um primeiro teste faz com que aqueles que participaram aprendam alguma coisa de seu mecanismo. Assim, os participantes terão melhor desempenho quando o teste for repetido tendo X ocorrido ou não;
- v) regressão estatística: num experimento que utiliza análises anteriores e posteriores a um tratamento, a variação captada no pós-teste não necessariamente representa o efeito de uma variável X específica. Indivíduos que estão muito afastados da média tendem a se reaproximar da mesma a despeito de qualquer espécie de tratamento;
- vi) vieses: serão produzidos em função das diferenças entre o grupo tratado e o de controle devido ao processo de seleção do último. Refere-se à não equivalência entre os dois grupos que estão sendo comparados;
- vii) mortalidade experimental: relaciona-se com a diferença entre as taxas de desistência associadas aos participantes de ambos os grupos promovendo a não equivalência de ambos;
- viii) interação entre seleção e maturação: ocorre em alguns desenhos *quasi*-experimentais onde múltiplos grupos são utilizados podendo ser confundidos com os efeitos da variável experimental.

Já a validade externa, de acordo com o mesmo autor, pode sofrer quatro principais ameaças, que são:

- i) reação a testagem: diz respeito ao processo de pré-testagem no qual foi submetido a amostra. Este pode alterar a sensibilidade da mesma quando exposta a X tornando os resultados não representativos para o conjunto da população não testada;
- ii) os efeitos de interação entre os vieses de seleção e a variável experimental;
- iii) efeitos de reação a arranjos experimentais: diz respeito à impossibilidade de generalizar resultados para indivíduos expostos a variável experimental em situações não-experimentais;
- iv) interferência de múltiplos tratamentos: quando a amostra é exposta a múltiplos tratamentos pode haver interferência entre eles, pois muitas vezes há memória do tratamento anterior;

## 2.1 Planejamento de experimentos

Seguindo a taxonomia proposta por CAMPBELL(1966), os desenhos experimentais podem ser divididos em três grandes grupos: pré-experimental; experimental e *quasi*-experimental. Para entender como as ameaças descritas na seção acima afetam os experimentos e quando e como estas podem ser controladas, uma breve apresentação dos grupos será feita utilizando alguns dos desenhos experimentais nelas contidos a título de ilustração do problema.

Para uma melhor compreensão dos desenhos experimentais a seguinte representação gráfica será utilizada:

- O, representa o processo de observação e mensuração da amostra;
- X, representa a ocorrência do tratamento;
- R, seleção aleatória do grupo ou randomização;
- cada linha representa as ocorrências relativas a um grupo;
- as ocorrências são apresentadas em ordem cronológica;

CAMPBELL(1966) apresenta três desenhos na categoria dos desenhos pré-experimentais. Eles são designados assim devido à pouca validade científica que apresentam e pela marcada falta de controle sobre vários fatores que se tornam hipóteses rivais a X na explicação das diferenças encontradas na amostra.

O caso mais extremo é o primeiro desenho em que um grupo é estudado uma só vez:

X O

O tratamento é feito e após sua ocorrência o grupo é observado. Sua pouca validade científica provém do fato de que nenhuma comparação é feita. Ao se descartar o processo de contraste na produção do conhecimento, o efeito de X é medido em relação ao que se esperava ou imaginava

da amostra antes do tratamento baseando-se no senso comum. Assim, o experimento trata de maneira desigual os lados que compara: observa e mede a amostra depois do tratamento com cuidado e zelo e tenta adivinhar o que teria ocorrido antes para poder medir o impacto de X. Este desenho apresenta quase todos os problemas citados na seção acima, pois possui inúmeras fontes de diferença além de X.

Outro desenho desta categoria pretende corrigir a falta de base comparativa através da inserção de um segundo grupo ou grupo de controle:

$$\begin{array}{c} X \ O_1 \\ \quad \quad O_2 \end{array}$$

Neste desenho o grupo tratado é exposto à intervenção e depois é submetido a observação enquanto o grupo de controle não sofre a exposição e é mensurado no mesmo momento que o tratado. O experimento supõe que todas as diferenças encontradas nas mensurações dos grupos devem ser atribuídas a X. Apesar de controlar algumas ameaças, neste desenho o viés de seleção se constitui como uma hipótese rival plausível, pois, nenhum procedimento foi adotado para garantir a equidade dos dois grupos. Assim, a despeito do tratamento existir, os grupos podem diferir per se já que nada garante que sejam equivalentes.

É a partir da necessidade intrínseca ao processo de produção de conhecimento, este tido como um processo de contraste e de comparação em detrimento da existência de um caráter absoluto do saber, que são propostos os verdadeiros desenhos experimentais. Todos os desenhos deste grupo possuem ao menos um grupo de controle onde a comparabilidade ou equivalência com o grupo dos tratados é devidamente assegurada, suprimindo desta forma as deficiências discutidas anteriormente nos desenhos pré-experimentais. Nesta categoria CAMPBELL(1966) aponta três desenhos consagrados pela literatura. Todos eles conseguem assegurar a validade interna do experimento apesar de terem alguns problemas em relação à generalização de seus resultados.

Para entender como tais desenhos conseguem garantir a equivalência entre o(s) grupo(s) dos tratados e o(s) grupo(s) de controle é preciso compreender como funciona o mecanismo de seleção utilizado para designar a participação nos grupos, a randomização. Este método consiste em selecionar aleatoriamente as pessoas a serem tratadas dentro da população elegível ao tratamento. Por população elegível entendem-se todas as pessoas que possuem a(s) característica(s) necessária(s) para poder atender ao tratamento, ou seja, que cumprem os

critérios do foco da intervenção. Assim, ao se estabelecer o critério de aleatoriedade para a participação no grupo dos tratados, criam-se dois grupos com distribuição equiprovável dos diversos valores de todos os outros fatores que determinam a variável sobre a qual se pretende medir o impacto (Y).

Em outras palavras, em função da seleção para a participação nos grupos não ter tido nenhuma orientação específica, espera-se que toda a heterogeneidade presente na amostra tenha sido representada igualmente tanto nos tratados como nos não tratados. Ou seja, o valor esperado de todos os outros fatores que determinam Y deve ser igual para os dois grupos, sendo assim, na média, a única diferença entre eles é a existência ou não da intervenção. Desta forma, o mecanismo de aleatoriedade permite a comparação direta dos resultados (Y) de ambos os grupos e a diferença entre estes passa a ser o impacto do tratamento. Esta técnica é extremamente vantajosa porque permite analisar o efeito da intervenção sem ter que estimar quais são todos os determinantes de Y, pois, ela atua sobre a média geral de todos eles.

Um exemplo deste grupo é o desenho:

$$RO_1 \quad X \quad O_2$$

$$RO_3 \quad O_4$$

Aqui, um grupo de tratados e um de controle é formado via randomização. Após a mensuração de ambos, aplica-se o tratamento no primeiro e depois é feita uma nova observação dos grupos. Neste desenho todas as ameaças à validade interna são controladas. Ao se garantir a igualdade entre os grupos, como exemplo, qualquer evento histórico ou efeito de maturação terá impacto igual no momento de mensuração dos grupos. Pode-se pressupor que ambos absorveram tais eventos de igual maneira, pois, uma vez que garantida a homogeneidade da dispersão dos parâmetros da regressão múltipla entre os grupos, espera-se que em conjunto, cada grupo tenha igual sensibilidade a qualquer evento idêntico que ocorra. Assim, quando se comparam as medidas, ou seja, quando contrastadas as diferenças  $O_2-O_1$  e  $O_4-O_3$ , o resultado será o efeito de X. Como a absorção dos outros eventos que afetam Y se deu da mesma maneira nos dois grupos, a diferença  $(O_2-O_1) - (O_4-O_3)$  provê o impacto de X. A inexistência de pré-teste exclui também a hipótese rival de testagem.

Entretanto, é preciso notar que os verdadeiros desenhos experimentais têm aplicação limitada. Isto ocorre devido ao fato de que nem sempre é possível que o pesquisador estabeleça a seleção aleatória, principalmente quando a análise se dá sobre dados observáveis e não sobre

experimentos controlados. Como exemplo, um programa governamental de saneamento básico não pode ser designado aleatoriamente, pois, deve atender a toda a população necessitada. Assim, seria muito difícil estimar seu efeito sobre, por exemplo, a saúde da população sem um grupo totalmente equiparável de controle. Desta forma o pesquisador deve encontrar alguma maneira alternativa para medir o impacto do tratamento. Esta categoria de desenhos é chamada de *quasi-experimental* e é marcada pela ausência de controle total do experimento pelo pesquisador, já que este não pode determinar o quando e o quem da exposição ao tratamento, mas somente os aspectos referentes à mensuração. CAMPBELL(1966) elenca uma série de tais desenhos que possuem um caráter específico ao tipo de área e problema que está sendo tratado. Na seção abaixo será discutida uma proposta de desenho *quasi-experimental* para avaliar o impacto da sindicalização na renda do trabalho no Brasil.

## **2.2 Uma proposta de desenho *quasi-experimental***

Apesar do desenho experimental descrito acima ser ideal para mensurar impactos, a impossibilidade de designar aleatoriamente indivíduos a participarem ou não dos sindicatos (estes tidos como tratamento) faz com que seja necessária a utilização de um desenho *quasi-experimental*. Para tal, o pesquisador deve identificar o grupo de trabalhadores sindicalizados e observar o comportamento da renda dentro do grupo. Em seguida ele deve se perguntar qual teria sido a renda dessas pessoas se o grupo não fosse sindicalizado e, através da comparação entre as duas situações, ele poderá medir o efeito da sindicalização na renda do trabalho.

Entretanto, tal tarefa se torna impossível na medida em que as pessoas não podem sofrer a intervenção e não a sofrer simultaneamente. Assim, haverá um problema de falta de dados no que se refere ao comportamento da renda quando o grupo sindicalizado não é sindicalizado. Se tais dados forem simplesmente substituídos por dados provenientes dos não sindicalizados haverá um viés na mensuração do impacto dado pela diferença entre a renda dos não participantes e a renda dos participantes caso não houvessem participado. Isto ocorre porque a renda do trabalhador depende não só da participação ou não no sindicato, mas também de inúmeros outros fatores como a escolaridade, a idade, o sexo, etc. Quando se selecionam não sindicalizados como base de comparação para os sindicalizados, a diferença entre eles no que diz respeito à renda pode não ser proveniente do tratamento, mas ser fruto de diferenças na idade ou em outros fatores. Assim, a discrepância entre os dados não estará medindo só a sindicalização

versus a não sindicalização, mas também qualquer outra diferença de condição entre os indivíduos a qual a renda do trabalho é sensível.

Uma alternativa seria a randomização, mas como já discutido anteriormente, esta se torna inviável. Desta forma, é preciso buscar outras maneiras de gerar dados bons o suficiente para substituírem as informações faltantes. Uma das alternativas é aplicar um questionário (mesmas pessoas) idêntico antes e depois da intervenção e utilizar os dados da primeira pesquisa como referência. A diferença dos  $Y$  entre os dois questionários proveria o impacto do tratamento. Entretanto, é preciso assegurar que não houve nenhuma variação nos demais determinantes de  $Y$  entre as duas situações, pois caso ocorra o contrário, a diferença entre os resultados das pesquisas mostrará algo além do impacto da intervenção. Além disso, um entrave para a utilização do recurso em discussão é a sua operacionalidade, pois tendo em vista que muitas vezes o pesquisador pretende avaliar uma intervenção que já se concretizou, ele deve se contentar com os dados disponíveis e, estes, muitas vezes não serão adequados para formar a base de comparação.

Outro método ao qual se pode recorrer para contornar o problema do viés é o pareamento. Tal técnica pretende buscar para cada indivíduo tratado um outro não tratado o mais semelhante possível ao primeiro, ou seja, encontrar o seu par. Supõe-se então que o par encontrado é a representação perfeita do indivíduo ao qual ele corresponde, com a única diferença de que sobre um incide a intervenção e sobre o outro não. Assim, o par estabelece o contrafactual do indivíduo tratado e a diferença de valores da variável escolhida entre ambos é o impacto do tratamento.

Para a que se possa utilizar o pareamento de forma bem sucedida é preciso atentar para uma série de questões. Antes de proceder com a busca pelo outro indivíduo é fundamental identificar adequadamente as variáveis que determinam  $Y$  (tais variáveis explicativas serão representadas por  $X$ , que doravante não mais se referirá ao tratamento). Problemas nesta tarefa resultarão em avaliações incorretas do tratamento e podem inclusive aumentar o viés da mensuração, uma vez que se julgará a semelhança entre os indivíduos sobre critérios parciais ou incorretos. Entretanto, quando se tem em mão uma boa regressão, o pesquisador poderá observar as características significantes dos indivíduos e pareá-los de acordo com a igualdade dos valores de  $X$ . Tendo feito isto é necessário comparar os valores que ambos apresentam para a variável escolhida e, a discrepância entre estes, será o impacto do tratamento. Note-se que é fundamental que para cada indivíduo tratado exista um correspondente não tratado para medir o efeito da intervenção.

Existem várias técnicas de pareamento e o *propensity score* é uma das propostas para tal feito. Este recurso é uma maneira específica de geração dos pares de indivíduos e se baseia na probabilidade de participação de cada indivíduo no tratamento como critério de igualação.

### 2.3 O pareamento e o *propensity score*

Tendo em vista que o caráter observacional do fenômeno em discussão, a saber, a ocorrência da sindicalização, se concretiza na esfera da decisão individual, não cabendo papel de designação exógena proveniente do pesquisador, torna-se inviável o recurso à randomização para a composição da amostra e de suas subcategorias. Desta maneira, uma nova preocupação deve ser levantada: como, então, se poderá assegurar a homogeneidade dos grupos que serão comparados, no que tange a tudo o mais que não o tratamento, uma vez que a exposição a este foi definida por critérios específicos ou resultou/sofreu influência de alguma conjunção singular de características existentes relevantes à questão em discussão? É a partir da tentativa de solucionar tal dificuldade, identificada pela literatura sob a consigna de “viés de seleção”, que o presente trabalho busca evidenciar a metodologia de pareamento e, em especial, uma maneira específica de realizá-la chamada *propensity score*. Como se buscará mostrar mais adiante, tal viés de seleção não pode ser totalmente eliminado em função da natureza da sua composição, entretanto a discussão em torno das possibilidades de minimizá-lo pode trazer a tona uma produção de dados significativamente mais verídica a cerca da realidade e, em seu bojo, todos os proveitos que dela derivam.

Uma vez identificado o problema, a existência do viés de seleção, torna-se necessário dispensar atenção para a sua definição precisa e para o reconhecimento das suas propriedades. Antes de prosseguir, cabe esclarecer melhor o significado do termo viés no contexto estatístico: ele relaciona-se ao conceito de tendenciosidade, ou seja, de um desvio sistemático do real valor do parâmetro que se deseja estimar. Assim, é preciso que o pesquisador busque algum algoritmo que possa corrigir tal disfunção recorrente do processo de estimação. O conceito de viés de seleção representa, então, um erro sistemático na estimação de um parâmetro, que pode assumir a forma de tratamento/intervenção, derivado do processo de seleção amostral.

Desta forma, a composição do grupo de controle se desviaria recorrentemente da plena representatividade do grupo de tratamento impossibilitando que se estabeleça com precisão o

contrafactual da situação em que os tratados não são submetidos ao tratamento. A equação a seguir denota o resultado  $Y$  para um indivíduo  $i$  qualquer:

$$Y_i = D Y_{1i} + (1-D) Y_{0i}$$

Onde  $D$  é uma variável *dummy* que ao assumir o valor 1 para o indivíduo  $i$  indica a ocorrência do tratamento e ao assumir o valor 0 representa a participação no grupo de controle e  $Y$  é o valor esperado para o indivíduo  $i$  da variável de interesse em que o subscrito 0 indica a não-participação no tratamento e o subscrito 1 a participação. Esta equação evidencia o problema do viés de seleção como um problema de *missing data* uma vez que se for observado  $Y_{0i}$  para o indivíduo  $i$  automaticamente será excluída a possibilidade de observar  $Y_{1i}$  para o mesmo indivíduo e vice-versa.

Assim, pode-se definir o viés como a diferença entre os tratados e os não-tratados antes da exposição ao tratamento, ou seja, a desigualdade inerente aos indivíduos que independente da realização do tratamento é captada na mensuração pós-tratamento. A formalização estatística do viés de seleção  $V$  é dada por:

$$V = E(Y_{0i} | D_i=1) - E(Y_{0i} | D_i=0)$$

Onde  $E$  representa a esperança matemática e  $|$  denota condicionalidade.

Em outras palavras, é a diferença no que diz respeito à determinação de  $Y$  entre aqueles que participaram do tratamento caso não o houvessem feito e aqueles indivíduos que de fato não foram tratados. A necessidade de buscar a representatividade perfeita (ou o mais próximo possível) dos indivíduos tratados caso não tivessem sido no grupo de controle decorre do fato de que  $E(Y_{0i} | D_i=1)$  não pode ser observado (assim como  $E(Y_{1i} | D_i=0)$  também não pode).

É através da procura por tal representatividade que surgem as técnicas de pareamento. Estas consistem em distintas proposições de regras para a formulação de algum algoritmo que permita encontrar no grupo de controle o indivíduo mais semelhante possível para cada indivíduo tratado. O pareamento (*matching*) busca estabelecer algum critério para mensurar o grau de igualdade entre os indivíduos baseando-se nas suas características observáveis, representadas pelo vetor  $X$  das co-variáveis que influenciam tanto a designação aos grupos como os resultados do tratamento. Dentro desta perspectiva, em cada par de indivíduos assim formado, uma vez que é garantida a igualdade das características observáveis que têm determinação em  $Y$ , há idêntica distribuição probabilística dos valores de  $Y$  (mesmo valor esperado de  $Y$ ). Ou seja, independentemente do indivíduo do par que seja selecionado para o

tratamento, espera-se igual valor para  $Y$ , assim sendo, ambos possuem a mesma resposta latente a ser desenvolvida pelo tratamento, pois as características sobre as quais o tratamento irá interagir e atuar para produzir o resultado são idênticas, fazendo com que o tratamento seja indiferente a pessoa específica que as porta.

Desta forma, dado  $X$ , os resultados potenciais independem do valor de  $D$ . A resposta ao tratamento condicionada ao vetor das co-variáveis é indiferente à participação ou não do indivíduo no mesmo. A partir daqui é que podem ser extraídas duas das hipóteses sobre as quais se assentam as técnicas de pareamento, nomeadamente a hipótese de identificação e a hipótese *da independência condicional*. Esta última implica na pressuposição de que os resultados são independentes da designação ao tratamento quando condicionados a  $X$  e, para que tal seja possível, deve vigorar a hipótese de identificação que estabelece ser viável identificar o vetor  $X$  das características observáveis que influenciam no processo de seleção e na resposta ao tratamento. De acordo com a hipótese da independência condicional, onde  $\perp$  indica independência, tem-se que:

$$(Y_{0i}, Y_{1i} \perp D_i | X)$$

Esta é a condição presente na randomização que permite a comparação direta dos resultados entre participantes e não-participantes, pois cada indivíduo da população tem a mesma probabilidade de receber o tratamento a despeito do vetor  $X$  de características. O pareamento busca simular tal característica dos mecanismos aleatórios e, a partir de tal consecução, pode tirar proveito das propriedades que derivam da independência condicional. Se tal condição é verdade então:

$$F(Y_{0i} | X_i, D_i=1) = F(Y_{0i} | X_i, D_i=0)$$

Onde  $F$  é uma função qualquer e se o cálculo das médias é possível então:

$$E(Y_{0i} | X_i, D_i=1) = E(Y_{0i} | X_i, D_i=0) = E(Y_{0i} | X_i)$$

De acordo com Rosenbaum e Rubin (1983), é chamado *escore de balanceamento* qualquer função que permita agrupar unidades do controle e do tratamento tal que a comparação direta dos resultados seja mais significativa. Ou seja, qualquer função  $B(X)$  tal que:

$$X \perp D | B(X)$$

Onde a forma com que  $X$  se associa com cada valor de  $B(X)$  independe da participação no tratamento, que é uma outra maneira de dizer que a distribuição condicional de  $X$  dado  $B(X)$  é

a mesma para  $D=0$  e  $D=1$ . Se a isto adicionarmos a condição que segue se terá o que os autores chamam de forte ignorabilidade no que se refere à designação ao tratamento:

$$0 < \text{pr}(D = 1 | X) < 1$$

Tal proposição se verdadeira implica em:

$$0 < \text{pr}(D = 1 | F(X)) < 1$$

Segundo os mesmos até os métodos usuais de ajuste multivariado (pareamento através de mais de uma variável) implicitamente realizam o ajuste através da estimativa de um escore de balanceamento escalar e  $B(X) = X$  seria a função mais refinada que é um escore de balanceamento enquanto que o *propensity score* ( $P(X) = \text{pr}(D = 1 | X)$ ) seria a função mais grosseira de  $X$  que também o é. Qualquer função  $B(X)$  para ser um escore de balanceamento deve então ser melhor que  $P(X)$  no sentido de que  $P(X) = F\{B(X)\}$  para qualquer função  $F$ . A vantagem da utilização de tal tipo de função como  $P(X)$  é a praticidade dada na composição do grupo de controle quando se utiliza um escore univariável como *proxy* de  $X$ , suprimindo as dificuldades acarretadas pela multidimensionalidade do vetor  $X$  das co-variáveis.

O outro pressuposto necessário ao funcionamento do pareamento e que evidencia categoricamente a vantagem da utilização de uma medida-resumo (escalar) é a hipótese do suporte comum. Esta estabelece que para cada conjunto de valores existentes do vetor  $X$  deve estar associado ao menos um indivíduo de cada grupo para tornar possível a mensuração do impacto do tratamento, também conhecida na literatura como efeito médio do tratamento sobre o tratado (average treatment effect on treated – ATT). Deve-se perceber que é mais fácil encontrar dois indivíduos idênticos sobre a base de uma medida univariável do que o contrário, uma vez que na prática os pesquisadores se deparam com uma base restrita de dados. O ATT é então obtido pela seguinte equação:

$$E(Y_{1i} - Y_{0i} | D_i=1, X) = E(Y_{1i} | D_i=1, X) - E(Y_{0i} | D_i=0, X)$$

A formulação acima indica que o efeito médio do tratamento sobre os tratados é o valor esperado da diferença entre o resultado estimado com o tratamento e sem o tratamento do iésimo indivíduo condicional a sua participação no grupo tratado e ao vetor  $X$ . Assumindo a hipótese da independência condicional é possível calcular o ATT através do segundo lado da igualdade, pois se pode estimar  $(Y_{0i} | D_i=1, X)$ , impossível de ser observado, através da estimativa de  $(Y_{0i} | D_i=0, X)$  dado que o condicionamento a  $X$  torna os resultados indiferentes ao valor da variável  $D$ . Assim a última equação pode ser re-escrita como segue:

$$E(Y_{1i} - Y_{0i} | X) = E(Y_{1i} | X) - E(Y_{0i} | X)$$

Por fim, os autores declaram que se a designação ao tratamento é fortemente ignorável dado  $X$  então, as diferenças nas médias dos dois grupos para cada valor de um escore de balanceamento é uma estimativa não viesada do efeito do tratamento para aquele valor e, como consequência, o pareamento, a subclassificação e o ajuste de covariância através de um escore de balanceamento podem produzir estimativas não viesadas do ATT. Este último seria obtido pela média das estimativas amostrais para cada valor do escore de balanceamento que produziria o balanceamento amostral em  $X$ .

Entretanto, tal formulação entra em contradição com aquela feita por Heckman no que diz respeito à possibilidade de produzir estimativas não viesadas do impacto do tratamento. Heckman ao discutir a natureza do viés de seleção afirma que tal viés pode ser desmembrado em três componentes, a saber:

- viés devido à falta de suporte comum;
- viés devido aos observáveis;
- viés devido aos não-observáveis;

Os primeiros dois componentes do viés podem ser eliminados através de uma rica base de dados que forneça o suporte comum e da identificação adequada das características observáveis que interferem seja na seleção seja no resultado. Entretanto, é a terceira dimensão do viés, a relação sistemática entre participação e resultados devido à existência de características não-observáveis, que torna o viés de seleção impossível de ser totalmente eliminado e torna o problema central da avaliação de tratamento em dados observacionais um problema de minimização do viés de seleção, dado que não se pode modelar sobre algo que não pode ser reconhecido através do processo de observação. Esta é a perspectiva que será adotada no presente trabalho.

Assim, trata-se agora de como encontrar o grupo de controle o mais próximo possível ao grupo dos tratados em termos de características observáveis. Como já mencionada, a técnica que será apresentada para tal é o *propensity score*, uma das concretizações específicas da metodologia conhecida na literatura como *matching* nos observáveis.

Seja  $P(X)$  a probabilidade de um indivíduo sofrer o tratamento condicional ao vetor  $X$  das co-variáveis, o *propensity score* busca então estabelecer o valor de tal probabilidade para cada indivíduo existente na base de dados e, através da comparação de tais valores, promover o

emparelhamento dos indivíduos. É importante notar que concretamente só existem duas situações distintas: a de participação no tratamento (valor 1) e a de não-participação (valor 0). Entretanto é possível construir uma medida contínua hipotética entre tais valores que reflita a contribuição da variação de status de cada característica relevante à designação do tratamento. Tal feito só é possível devido à análise da recorrência dos valores das variáveis que determinam à participação no tratamento, ou seja, daquelas variáveis que apresentam um conjunto de valores específicos ou uma tendência definida para cada valor de D. Assim, através da associação dos valores das variáveis (medidos antes do tratamento) que são mais comuns à situação onde  $D = 1$  em contraste com os valores mais comuns associados quando  $D = 0$  é que se constrói, através da ponderação, tal medida intermediária. Pois, determinado indivíduo pode apresentar uma combinação de características em que algumas variáveis tenham valores recorrentes ao grupo de participantes e outras não. Assim, a caracterização do indivíduo tratado ideal (aquele que apresenta os valores de todas as características relevantes a designação à participação no tratamento no espectro de valores recorrentes a situação  $D = 1$ ) e do não tratado ideal permite construir uma gradação da chance de participar no tratamento. Desta forma, em dados observacionais, as estimativas de  $P(X)$  são necessariamente posteriores ao tratamento.

Como  $P(X)$  deve ser um valor tal que  $0 < P(X) < 1$ , pois não pode existir probabilidade negativa e nem superior a 1 de participação e, ao mesmo tempo, uma variável independente, a sua estimativa pode ser feita através de uma regressão do tipo probit/logit dado que o modelo de regressão linear não comporta estas condições. Então, na regressão logística devem ser incluídas todas as variáveis presentes nos dados que podem influenciar na participação assim como todas aquelas que afetam a variável Y de interesse para poder obter os valores preditos de  $P(X)$ .

De acordo com a hipótese da independência condicional se a participação no tratamento é fortemente ignorável dado X então ela também será para qualquer função de X como  $P(X)$ :

$$(Y_{0i}, Y_{1i} \perp D_i | X) \text{ e } E(Y_{0i} | X_i, D_i=1) = E(Y_{0i} | X_i, D_i=0)$$

então,

$$(Y_{0i}, Y_{1i} \perp D_i | P(X)) \text{ e } E(Y_{0i} | P(X), D_i=1) = E(Y_{0i} | P(X), D_i=0)$$

e o ATT dado pela fórmula abaixo já apresentada anteriormente

$$E(Y_{1i} - Y_{0i} | D_i=1, X) = E(Y_{1i} | D_i=1, X) - E(Y_{0i} | D_i=0, X)$$

que pode então ser calculado como

$$E(Y_{1i} - Y_{0i} | D_i=1, P(X)) = E(Y_{1i} | D_i=1, P(X)) - E(Y_{0i} | D_i=0, P(X))$$

também expresso da seguinte maneira

$$E(Y_{1i} - Y_{0i} | P(X)) = E(Y_{1i} | P(X)) - E(Y_{0i} | P(X))$$

Depois da obtenção dos valores de  $P(X)$  são calculadas as diferenças para cada indivíduo tratado entre a sua probabilidade de participação predita e as probabilidades de todos os outros indivíduos não-tratados da base de dados. Os  $n$  indivíduos com as  $k$  menores diferenças são eleitos para o emparelhamento e passam assim a estabelecer o contrafactual da situação em que os tratados não seriam tratados possibilitando o cálculo do ATT.

O presente trabalho adotou regras específicas para realizar o pareamento dos escores de propensão. Entretanto é importante evidenciar que existem outros critérios possíveis para realizá-lo. Para a análise de caso será utilizado  $k=1$ . Quando houver mais de um indivíduo portando a menor diferença encontrada para parear determinada observação será selecionado somente um deles, de maneira aleatória, para constituir o contrafactual da observação em questão. Ao se utilizar uma observação do controle na formação de um par a mesma será descartada como candidata a parear qualquer outra observação.

### Capítulo 3: O caso da sindicalização sobre o rendimento do trabalho

A partir de uma amostra de 170.418 indivíduos ocupados remunerados na semana, a PNAD indicava que, em 2006, o Brasil apresentava um universo de 14,7 milhões de sindicalizados, que auferiam um rendimento médio 68% superior ao dos 63,7 milhões de não sindicalizados (Tabela 1). Destaca-se ainda a forte assimetria da distribuição de rendimentos em ambas as populações, que apresentam valores médios substancialmente superiores ao valor mediano.

Tabela 1 – Estatísticas descritivas para rendimento do trabalho principal segundo a sindicalização<sup>(1)</sup> – Brasil 2006

	<b>Sindicalizados</b>	<b>Não Sindicalizados</b>
<b>N</b>	31.487	138.931
<b>N</b>	14.676.399	63.694.356
<b>Média</b>	1.263,63	751,82
<b>Mediana</b>	700,00	440,00
<b>Desvio Padrão</b>	1.862,31	1.370,65

<sup>(1)</sup> Excluídos ocupados com rendimentos nulos ou não declarados

Fonte: PNAD, microdados, IBGE.

Os dados acima podem ter melhor compreensão quando analisados a luz das diferentes características ou tendências, presentes nos grupos de indivíduos sindicalizados e não sindicalizados, que de alguma maneira são relevantes para a determinação da renda. Ou seja, para além da primeira pressuposição de que o fator sindicalização tem impacto no rendimento do trabalho é preciso averiguar entre os grupos se há alguma assimetria significativa nos demais aspectos que influenciam a composição da renda para que não sejam atribuídas enganosamente à sindicalização as divergências nas medidas-resumo. Desta forma, parece útil à tarefa proposta pelo presente trabalho construir aqui o perfil dos sindicalizados e dos não sindicalizados para que seja possível agrupar os fatores que conjuntamente geram tamanha desigualdade nos rendimentos de ambos os grupos. Posteriormente se tentará destrinchar a influência isolada de cada um deles. Abaixo seguem os dados (Tabela 2) que servirão de base para a construção de tais perfis:

Tabela 2 - Valores médios segundo características sócio-econômicas e sindicalização

	<b>Total</b>	<b>Não Sindicalizados</b>	<b>Sindicalizados</b>
Jornada	41,24	41,08	41,92
No	7%	8%	5%
Ne	24%	24%	25%
Sp	24%	24%	24%
se_sp	21%	22%	19%
Co	8%	8%	7%
Su	16%	15%	20%
Rm	33%	33%	32%
Urbano	86%	87%	83%
Anosest	7,95	7,72	8,94
Idade	36,75	35,93	40,31
Masc	60%	58%	65%
Branco	52%	51%	58%
agricultura	12%	11%	17%
Industria	16%	15%	20%
construcao	7%	8%	3%
Comrepar	19%	20%	12%
Alojalim	4%	4%	2%
Transcom	5%	5%	7%
Admpub	6%	5%	8%
Edsauss	10%	8%	16%
Servdom	9%	10%	1%
Outrativ	13%	13%	13%
Dirigente	6%	5%	7%
professional	7%	6%	13%
Tecnico	8%	7%	11%

Services	41%	44%	30%
Producao	25%	25%	22%
Forcasarm	1%	1%	1%
Agricola	12%	11%	17%
Outrocp	0%	0%	0%
Empcart	38%	34%	54%
Empscart	26%	30%	8%
Clemil	7%	5%	16%
Cntprop	24%	25%	18%
Empr	5%	5%	5%
Outrpos	0%	0%	0%

Fonte: PNAD, microdados, IBGE. Elaboração dos autores

Em média percebe-se que os sindicalizados possuem um nível de instrução superior (8,94 anos de estudo) aos não sindicalizados (7,72 anos de estudo) ao mesmo tempo em que apresentam idade mais avançada. Até o momento, mesmo com estas poucas informações é possível construir uma série de arranjos explicativos distintos para compreender as diferenças encontradas na renda do trabalho. Por um lado, pode-se dizer que as variáveis anos de estudo e renda do trabalho possuem um comportamento diretamente proporcional e que talvez a sindicalização não seja o fator que promove a diferenciação da renda, mas que aqueles que estudam mais além de ganhar mais também possuem maior propensão a sindicalização. Ou então, outra hipótese plausível seria a de que como os sindicalizados em geral são mais velhos, eles teriam tido mais tempo hábil para adquirir uma quantidade superior de anos de estudo e que, então, a chave para entender a renda mais alta de tal grupo estaria contida no intervalo de razões existentes para que as pessoas em uma faixa etária mais alta resolvessem se sindicalizar mais do que as outras. Poderia ser em decorrência de que estes, como tiveram mais anos de educação formal, tenham sido levados a concluir que a melhor opção seria a sindicalização ou, pelo contrário, que a educação formal em nada influa ou mesmo faça com que as tendências anti-coletivas sejam exacerbadas e que tal diferencial seja oriundo de algum evento histórico determinado vivido exclusivamente por tal geração de forma que esta seja mais propensa por isto a se sindicalizar.

O que se pretende destacar com este exercício de abstração é que tal tipo de análise enquanto instrumento de acepção da realidade possui fortes limitações. Ela só é uma explicação averiguável enquanto o conjunto dos fatores, não se pode identificar a contribuição de cada fator em específico e tampouco determinar qualquer cláusula causal entre eles. Uma vez que o objetivo aqui é compreender a relação entre renda do trabalho e sindicalização a construção do perfil do sindicalizado será meramente uma tarefa de caráter descritivo e não explicativo no que tange a questão. Neste estágio de análise, sem o aporte dos conhecimentos trazidos pelo estudo histórico, o tipo de verdade produtível só tem caráter explicativo enquanto o agregado dos fatores.

A par das restrições impostas pelo instrumental utilizado até aqui, não deixa de ser útil a construção do perfil dos sindicalizados como uma primeira aproximação do problema. Além das já mencionadas características pode-se perceber que, do ponto de vista da dispersão geográfica, a região norte e sudeste sem a presença do estado de São Paulo têm proporcionalmente uma menor concentração de sindicalizados em relação aos não sindicalizados como contrapartida da predominância dos primeiros na região sul. Ainda na perspectiva geográfica nota-se que os sindicalizados, a despeito de serem compostos majoritariamente por indivíduos advindos das zonas urbanas, têm maior incidência nas áreas rurais do que os não sindicalizados.

Apesar das predominâncias de homens e de brancos serem características gerais da amostra, dentre os sindicalizados estas tendências são ainda mais acentuadas. Quando comparados aos não sindicalizados, aqueles adeptos a sindicalização têm maior concentração relativa nas seguintes setores de atividade: agricultura, indústria, transações comerciais, administração pública e educação e saúde. E, do ponto de vista da ocupação o mesmo ocorre com os dirigentes, os profissionais e os técnicos tendo como contrapartida fundamentalmente o impacto nos serviços.

Por fim, pode-se dizer que mais da metade dos sindicalizados têm carteira assinada enquanto que o restante tende a ser funcionário público estatutário ou militar ou trabalhador por conta própria. É notável também a grande concentração dos não sindicalizados entre os ocupados sem carteira quando comparados com aqueles que aderiram à sindicalização. Assim, o sindicalizado é mais qualificado e mais velho, ocupa empregos formais e tende a ser branco e do sexo masculino.

Dadas as tendências identificadas na população dos sindicalizados, cabe agora tentar relacioná-las com a renda para entender como elas a impactam e assim poder isolar o efeito do sindicato na mesma. (Nota: se não houvesse tendências divergentes nas características das populações de sindicalizados e não sindicalizados o efeito da sindicalização poderia ser atribuído pela comparação direta das medidas resumo). Deveria-se, portanto, considerar outros fatores sócio-econômicos que possam estar afetando a renda do trabalho, tais como escolaridade, idade, sexo, entre outros. Em outras palavras, a magnitude da diferença observada na Tabela (1) não seria devida exclusivamente à sindicalização ou não do indivíduo, mas também a fatores exógenos também associados à decisão de sindicalização do indivíduo.

Um procedimento comum neste tipo de análise é elaborar um modelo de regressão múltipla para mensurar o efeito isolado de inúmeras características sócio-econômicas, inclusive sindicalização, sobre o rendimento médio do trabalho. Esse modelo poderia ser dado por:

$$\ln(Renda) = \alpha + \lambda \text{Sindicato} + \sum_i^k \beta_i X_k + e \quad (1)$$

Onde:

$\ln(Renda)$ : logaritmo neperiano da renda do trabalho principal;

$Sindicato$ : 1 se sindicalizado; 0 caso contrário;

$X_k$ : k-ésima variável independente;

$e$ : resíduos aleatórios e independentes;

Cada coeficiente  $\beta_j$  representaria, dessa forma, a diferença no logaritmo do rendimento hora do trabalho na ocupação principal associada a uma diferença na variável  $X_j$ . Em outras palavras, quando  $X_j$  é uma variável contínua, a diferença percentual na renda associada a uma diferença unitária em  $X_j$  será equivalente a  $100\beta_j\%$ . Quando  $X_j$  for uma variável dicotômica, que assume apenas valores 0 e 1, a diferença percentual na renda associada à dicotomia estabelecida será equivalente a  $100[e^{\beta_j}-1]\%$  (HALVORSEN; PALMQUIST, 1980).

Seguindo as sugestões de HOFFMANN (2000) para uma função de rendimentos foram consideradas, além da sindicalização, as seguintes variáveis independentes:

- i) Jornada de trabalho (*jornada*): horas trabalhadas na semana;

- ii) Região geográfica: cinco variáveis binárias (*ne, sp, se\_sp, co, su*) para identificar seis regiões geográficas (Nordeste; São Paulo; Sudeste sem São Paulo; Centro-Oeste; Sul; e Norte, utilizado como referência);
- iii) Região Metropolitana (*rm*): 1 se for reside em Região Metropolitana, 0 para demais localidades;
- iv) Situação censitária (*urbano*): 1 se reside em área urbana, 0 para rural;
- v) Escolaridade (*anosest*): anos completos de escolaridade;
- vi) Idade: anos de idade (*idade*) e quadrado dos anos de idade (*idade2*) para considerar rendimentos decrescentes a partir de determinada idade;
- vii) Sexo (*masc*): 1 se do sexo masculino, 0 se feminino;
- viii) Cor ou raça (*branco*): 1 se da cor branca ou amarela, 0 caso contrário;
- ix) Setor de atividade: nove variáveis binárias (*industria, construcao, comrepar, alojalin, transcom, admpub, edsauss, servdom, outrativ*) para considerar dez agrupamentos de atividade econômica do empreendimento onde a pessoa trabalhava (atividades agrícolas, utilizada como referência da análise; indústria; construção; comércio e reparação; alojamento e alimentação; transporte armazenagem e comunicação; administração pública; educação, saúde e serviços sociais; serviço doméstico; e outras atividades);
- x) Ocupação: sete variáveis binárias (*dirigente, profissional, tecnico, servicos, producao, forcasarm, outrocp*) para identificar oito agrupamentos ocupacionais (trabalhadores agrícolas, utilizado como referência; dirigentes em geral; profissionais das ciências e das artes; técnico de nível médio; trabalhadores de serviços; trabalhadores da produção de bens e serviços, de reparação e manutenção; membros das forças armadas; e outras ocupações);
- xi) Posição na ocupação: cinco variáveis binárias (*empcart, clemil, cntprop, empr, outrpos*) para considerar seis posições na ocupação (empregado sem carteira, utilizado como referência; empregado com carteira; funcionário público estatutário ou militar; conta-própria; empregador; e outras posições);

As estimativas de mínimos quadrados ordinários (MQO) para os coeficientes do modelo (Tabela 3) são, em sua grande maioria, significativas e mostram coerência com os resultados

apresentados na literatura (HOFFMANN, 2000). Para este trabalho, o fato relevante seria destacar a significância da sindicalização na determinação da renda do trabalho. Como sugere a interpretação das estimativas, o rendimento médio dos sindicalizados seria, em média, 12% superior ao dos não sindicalizados, independente dos demais fatores sócio-econômicos também associados à determinação da renda do trabalho.

Tabela 3 – Estimativas de Mínimos Quadrados Ordinários para equação (1)

Variável	$\hat{\beta}$	t	P
intercepto	2,9082	199,77	***
sindicato	0,1133	26,91	***
jornada	0,0154	124,26	***
ne	-0,3337	-51,05	***
sp	0,0963	14,07	***
se_sp	-0,0630	-9,31	***
co	0,0981	12,19	***
su	0,0234	3,26	0,0011
rm	0,1186	33,7	***
urbano	0,0926	16,65	***
anosest	0,0713	144,27	***
idade	0,0635	102,62	***
idade2	-0,0006	-81,11	***
masc	0,3332	87,62	***
branco	0,1167	33,91	***
industria	0,0950	5,02	***
construcao	0,1749	8,88	***
comrepar	0,0932	4,89	***
alojalim	0,0516	2,53	0,0114
transcom	0,2485	12,46	***
admpub	0,3033	14,82	***
edsauss	0,0638	3,23	0,0012

servdom	-0,0162	-0,82	0,4138
outrativ	0,1552	8,06	***
dirigente	0,5059	25,41	***
profissional	0,5671	28,35	***
tecnico	0,3150	16,02	***
servicos	0,0024	0,12	0,9015
producao	0,0337	1,78	0,0757
forcasarm	0,2424	9,2	***
outrocp	-0,5403	-2,97	0,0029
empcart	0,2806	64,08	***
clemil	0,3616	43,04	***
entprop	-0,0261	-5,32	***
empr	0,6152	65,06	***
outrpos	0,2159	0,71	0,4793
<hr/>			
F	7149,73	(***)	
R <sup>2</sup>	0,5964		

(\*\*\*) Significância a 0,01%

Fonte: PNAD, microdados, IBGE. Elaboração dos autores.

O que as estimativas a cima tentam realizar é a mensuração da contribuição isolada de cada determinante da renda. Assim, mantendo k-1 variáveis explicativas constantes permite-se alternadamente que cada variável assumam diferentes valores ao mesmo tempo em que são observadas e registradas as diferenças provocadas em Y por tais variações. De maneira geral esta é a forma em que o MQO pode estimar a contribuição de cada variável explicativa para a determinação de Y.

Observando as estimativas da tabela 3 e fazendo uma descrição que procure destacar os dados de uma maneira mais geral e que pressuponha para cada variável discutida a independência em relação às demais, pode-se perceber que os homens têm um rendimento médio 39,51% superior ao das mulheres, os brancos 12% superior aos dos não-brancos e que cada ano

de estudo adicional eleva a renda média em 7,13%. Aqueles que têm seu rendimento principal oriundo do trabalho e pertencem a áreas urbanas têm em média um acréscimo de 9,75% na sua renda quando comparados aos habitantes das regiões rurais. De maneira análoga, nota-se uma elevação de 12,64% no rendimento médio daqueles que vivem nas regiões metropolitanas em relação aos moradores de demais localidades. Para cada hora adicional na jornada de trabalho semanal está associado um aumento médio de 1,54% na renda.

Do ponto de vista da localização geográfica, tendo os habitantes da região Norte como base, as maiores discrepâncias na renda média são encontradas na região Nordeste, que apresenta uma redução de 28,39% para a renda média de seus habitantes, seguida da região Centro-Oeste, onde há um aumento de 10,30%, acréscimo este muito próximo ao valor registrado para os moradores do Estado de São Paulo (10,08%).

Para uma análise que se ancore no setor de atividade onde está localizado o trabalhador e que tenha como referência as atividades agrícolas, os dados indicam que os maiores rendimentos médios são encontrados na administração pública, nas atividades de transporte armazenagem e comunicação e na construção, onde tal ordenação dos setores de atividade corresponde ao critério de rendimentos médios decrescentes. Assim, foram estimados aumentos de 35,39%, 28,27% e 19,12% respectivamente. Os dados também apontam que a renda média dos trabalhadores localizados nos setores de atividade de educação, saúde e serviços sociais e de serviço doméstico não é estatisticamente diferente da dos trabalhadores em atividades agrícolas.

No que diz respeito aos grupamentos ocupacionais onde se distribuem os trabalhadores, nota-se que são os dirigentes em geral e os profissionais das ciências e das artes que auferem os maiores rendimentos médios. Em relação à renda média dos trabalhadores agrícolas, o rendimento médio dos primeiros se mostra 65,86% e 76,30% superior respectivamente. Por fim, quando considerada a posição na ocupação com referência nos empregados sem carteira, os maiores rendimentos médios foram encontrados entre os empregadores e os funcionários públicos, estatutários ou militares, rendimentos estes que seriam superiores ao dos trabalhadores agrícolas em 84,97% e 43,62% respectivamente.

Entretanto, tais estimativas devem ser vistas com grande cautela, pois um dos pressupostos dessa análise é a independência entre o resíduo  $e$  e as variáveis independentes do modelo, inclusive *sindicato*. Em outras palavras, podem existir fatores que afetam a seleção na participação do sindicato, que ocorre via auto-seleção e não através de um mecanismo aleatório,

e que também influenciam a renda do trabalho. Considerando  $D$  como uma variável binária que representa a ocorrência da sindicalização caso seja 1 e a sua não ocorrência caso seja 0 e que esta variável pode ser estimada através de algum tipo específico de regressão, existem duas possibilidades de correlação com os resíduos da regressão da renda: a correlação dos resíduos da regressão de  $D$  com  $e$  e a correlação das variáveis explicativas de  $D$  com o último. Assim sendo,  $D$  pode estar rompendo tal pressuposto de duas maneiras distintas. Caso existam fatores, presentes nos resíduos da regressão da renda, que afetem simultaneamente a renda e a decisão de sindicalização dos indivíduos através das variáveis explicativas de  $D$  e/ou dos seus resíduos, as estimativas de MQO seriam inconsistentes, ou seja, viesadas mesmo para amostras grandes. A primeira forma do problema pode ser resolvida pela inclusão das variáveis explicativas de  $D$  que afetam a renda na regressão desta, porém, se existe correlação entre os resíduos, esta persiste independentemente do recurso a tal procedimento. Por exemplo, atitudes profissionais do indivíduo no ambiente de trabalho, fator não observável no modelo, poderiam influenciar seu rendimento  $e$ , da mesma forma, sua decisão em relação à sindicalização.

Para avaliar o impacto da sindicalização sobre a renda do trabalho seria necessário identificar qual a renda do trabalhador sindicalizado caso este não fosse sindicalizado. Tendo em vista que tal situação não pode ser verificada, uma das alternativas para contornar este problema é a aplicação de alguma técnica de *matching*, no caso do presente trabalho a técnica selecionada é o *propensity score*. A par das limitações de tal categoria de técnicas, a saber: tais técnicas somente lidam com as diferenças observáveis, é que prosseguirá a presente investigação. Assim, ainda existe uma heterogeneidade latente devido a fatores não observáveis que pode resultar em uma estimativa viesada do impacto do tratamento. A técnica de *propensity score* procura dividir a amostra em dois grupos com valores iguais ou semelhantes de  $X$  e compara as médias condicionais para sindicalizados e não sindicalizados. O *propensity score* procura realizar a comparação entre valores nos grupos de tratamento e controle a partir do pareamento de observações relativamente homogêneas. Eliminar-se-ia, desta forma, observações que não apresentam características homogêneas nos dois grupos e poderiam estar causando algum viés de seleção.

Para aplicar o *propensity score*, o primeiro passo é associar para cada indivíduo da amostra uma probabilidade predita de participação no tratamento, o que pode ser feito com o uso

da regressão logística. A regressão logística relacionando o *odds ratio* da propensão à sindicalização às variáveis explanatórias seria dada por:

$$\ln\left(\frac{Sindicato_i}{1 - Sindicato_i}\right) = \alpha + \sum_i^k \beta_i X_k + e_i \quad (2)$$

Foram utilizadas as mesmas variáveis explicativas da equação de rendimentos para determinar a propensão à sindicalização. A partir das estimativas de máxima verossimilhança dos coeficientes apresentadas na Tabela 4 pode-se interpretar como os diferentes fatores associam-se à condição de sindicalização do trabalhador. Por exemplo, a sindicalização é maior entre ocupados do sexo masculino, brancos, na região Sul e Nordeste, independente dos demais fatores. Além da interpretação dos coeficientes, que oferece um rico material para futuras análises sobre o tema, as estimativas foram ainda utilizadas para estimar a propensão à sindicalização de cada ocupado, seja ele sindicalizado ou não. A partir de (2), pode-se estimar a propensão à sindicalização por:

$$Pr ob(Sindicato_i = 1) = \frac{e^{\alpha + \sum_{i=1}^k \beta_i X_i}}{1 + e^{\alpha + \sum_{i=1}^k \beta_i X_i}} \quad (3)$$

Tabela 4 – Estimativas de Máxima Verossimilhança para equação (2)

z	Estimativa	X2	p
Intercepto	-4,3023	1612132	***
jornada	0,0005	331	***
ne	0,3956	74050	***
sp	0,0534	1242	***
se_sp	0,0035	5	0,0192
co	0,0624	1234	***
su	0,3763	59176	***
rm	-0,0437	3568	***
urbano	-0,3927	123744	***

anosest	0,0665	413806	***
idade	0,0759	308013	***
idade2	-0,0005	105484	***
masc	0,1754	53056	***
branco	0,0140	381	***
industria	0,0215	40	***
construcao	-0,8009	45312	***
comrepar	-0,5076	21193	***
alojalim	-0,5605	20430	***
transcom	0,1876	2706	***
admpub	-0,3020	6596	***
edsauss	-0,1739	2357	***
servdom	-1,9726	194202	***
outrativ	-0,1837	2747	***
dirigente	-0,6450	32561	***
profissional	-0,4770	17462	***
tecnico	-0,6463	33092	***
servicos	-0,8329	58581	***
producao	-0,7916	53407	***
forcasarm	-1,7935	121357	***
outrocp	-0,6116	393	***
empcart	1,7748	2323700	***
clmil	2,0077	1475357	***
cntprop	0,4864	144988	***
empr	0,7854	170278	***
outrpos	1,1317	213	***

---

Max-rescaled R2      0,2246

(\*\*\*) Significância a 0,01%

Fonte: PNAD, microdados, IBGE. Elaboração dos autores.

As estimativas da Tabela 4 referem-se ao *odds ratio*, ou seja, elas devem ser interpretadas como medidas que se relacionam à probabilidade de sucesso sobre a probabilidade de fracasso. Ao definir o sucesso como a ocorrência da sindicalização, cada coeficiente  $\beta_j$  representaria, dessa forma, a diferença no logaritmo do *odds ratio* associada a uma diferença na variável  $X_j$ . Em outras palavras, quando  $X_j$  é uma variável contínua, a diferença percentual no *odds ratio* associada a uma diferença unitária em  $X_j$  será equivalente a  $100\beta_j\%$ . Quando  $X_j$  for uma variável dicotômica, que assume apenas valores 0 e 1, a diferença percentual no *odds ratio* associada à dicotomia estabelecida será equivalente a  $100[e^{\beta_j}-1]\%$  (HALVORSEN; PALMQUIST, 1980).

Desta forma, ao analisar tais estimativas, pode-se perceber quais as características que mais impactam a probabilidade de ser sindicalizado sobre a de não ser independentemente dos demais fatores. De acordo com as regiões geográficas (referenciadas na região norte), os trabalhadores das regiões nordeste e sul apresentam as maiores probabilidades de serem sindicalizados com 48,53% e 45,69% a mais de chances respectivamente. Entretanto, ser um trabalhador urbano impacta negativamente tais chances em 32,48%.

Cada ano adicional de estudo se reflete em um aumento de 6,65% do *odds ratio*, da mesma forma, cada ano a mais na idade afeta positivamente em 7,59%. Já o fato do trabalhador ser do sexo masculino eleva tal relação de probabilidades em 19,17%.

Do ponto de vista do setor de atividade em que se localiza o indivíduo e tendo como base as atividades agrícolas, as menores chances de ser sindicalizado encontram-se nas atividades de serviço doméstico, de construção, de alojamento e alimentação e de comércio e reparação. Respectivamente tais chances são reduzidas em 86,09%, 55,11%, 42,91%, 39,81%. Entretanto deve ser destacado que o único setor de atividade que tem as suas chances aumentadas é o de transporte, armazenagem e comunicação com 20,64% a mais do que o agrícola.

Em relação à ocupação, pode ser dito que os trabalhadores agrícolas são aqueles que apresentam o maior valor do *odds ratio* visto que os demais grupamentos ocupacionais têm efeito redutor em tal valor quando referenciados aos primeiros. Destacam-se os membros das forças armadas com -83,36%, os trabalhadores de serviços com -56,52% e os trabalhadores da produção de bens e serviços com - 54,69%. Por fim, pode-se dizer que a posição na ocupação é o fator que mais altera a chance de ser sindicalizado em relação à chance de não ser. Tendo como base da análise os empregados sem carteira, os funcionários públicos, estatutários ou militares apresentam 644,62% a mais de chances de serem sindicalizados, os empregados com carteira

489,91%, outras posições 210,09%, os empregadores 119,33% e os trabalhadores por conta-própria 62,65%.

A partir de tais estimativas, observações nos grupos de sindicalizados e não sindicalizados com propensões à sindicalização semelhantes foram pareadas. Considerou-se uma aproximação na terceira casa decimal para equivalência das propensões  $Prob(Sindicato_i = 1)$ . Uma precisão maior, com maior número de casas decimais, resultaria em poucos pares e uma precisão menor sujeitaria o pareamento a um maior viés de seleção.

Das 31.487 observações presentes no arquivo de caso (sindicalizados) e 138.931 observações presentes no arquivo de controle (não sindicalizados), foram pareadas 30.915 observações, ou seja, 98% das observações onde houve a ocorrência da sindicalização. Para o presente estudo pôde, então, ser constatado que do ponto de vista quantitativo o método foi eficiente no que diz respeito ao aproveitamento das informações fornecidas pela a base de dados. Isso significa que cada um dos 30.915 indivíduos do grupo de sindicalizados apresenta um par no grupo de não sindicalizados com características sócio-econômicas semelhantes, com a única diferença de um ser e o outro não ser sindicalizado.

Tabela 5 – Valores médios das observações pareadas segundo características sócio-econômicas e sindicalização

	<b>Não Sindicalizados</b>	<b>Sindicalizados</b>
jornada	42,38623724	41,97535351
no	5,658563%	5,518001%
ne	23,971860%	24,715846%
sp	23,983302%	23,790886%
se_sp	20,090415%	19,485053%
co	6,905220%	6,794794%
su	19,390639%	19,695420%
rm	32,113642%	32,015363%
urbano	82,692526%	82,788307%
anosest	8,83778165	8,88978570
idade	39,88769285	40,20144780

idade2	1745,66345456	1774,96149759
masc	64,780532%	64,595645%
branco	57,701908%	57,534704%
agricultura	17,520307%	17,491796%
industria	20,993728%	20,391607%
construcao	2,618591%	3,030545%
comrepar	12,331517%	12,327083%
alojalim	2,216961%	2,186099%
transcom	6,885100%	6,654223%
admpub	7,881550%	8,030824%
edsauss	15,279705%	15,874504%
servdom	0,995930%	0,891094%
outrativ	13,276612%	13,122225%
dirigente	7,280807%	7,141192%
profissional	11,323902%	12,282604%
tecnico	10,813721%	10,841018%
servicos	30,256869%	29,788487%
producao	22,362426%	21,902699%
forcasarm	0,599265%	0,625114%
agricola	17,346423%	17,406955%
outrocp	0,016586%	0,011931%
empcart	57,372602%	53,781165%
empscart	6,759018%	7,608147%
clmil	13,844599%	15,293391%
cntprop	16,509554%	18,069734%
empr	5,514228%	5,246052%
outros	0,000000%	0,001511%

Fonte: PNAD, microdados, IBGE. Elaboração dos autores

Os dados da Tabela 5 mostram os valores médios de cada variável da regressão de propensão à sindicalização para ambos os grupos de tratados e de controle. A partir do

pareamento dos valores  $Prob(Sindicato_i = 1)$  como uma medida univariável das características X que influenciam a participação no tratamento e a determinação da renda, pode-se notar a equalização produzida nos valores médios dos diferentes Xi para os dois grupos. Assim, foi observado, por exemplo, que a porcentagem de indivíduos brancos antes do pareamento era de 51% entre os não sindicalizados e de 58% entre os sindicalizados. Tal diferencial gera impactos distintos na determinação da renda entre os grupos fazendo com que tais efeitos possam ser confundidos e mal interpretados como sendo produto da sindicalização. Após o pareamento os novos valores encontrados foram 57,70% e 57,53% respectivamente garantindo que tal característica afete a determinação da renda dos não-sindicalizados e dos sindicalizados de maneira homogênea. Desta forma, tal fenômeno de equalização foi verificado para as demais características observáveis do modelo tornando virtualmente idêntico o “indivíduo médio” de ambos os grupos no que se refere às mesmas. Assim sendo, a comparabilidade direta entre os grupos pareados torna-se possível permitindo identificar o efeito da sindicalização.

As principais estatísticas descritivas para cada um dos grupos são apresentadas na Tabela 6 e permitem identificar, por exemplo, que o rendimento médio dos sindicalizados é 18% superior à dos não sindicalizados, e não 12% superior como sugeriu o modelo (1). A diferença média mostrou-se ainda significativa a 0,01% quando realizado um teste para amostras dependentes.

Tabela 6 – Estatísticas descritivas para rendimento do trabalho principal segundo sindicalização<sup>(1)</sup> – Brasil 2006

	<b>Sindicalizados</b>	<b>Não Sindicalizados</b>
<b>n</b>	30.915	30.915
<b>Média</b>	1.251,33	1.056,80
<b>Mediana</b>	700,00	600,00
<b>Desvio Padrão</b>	1.879,16	1.877,79

<sup>(1)</sup> Exclusive ocupados com rendimentos nulos ou não declarados

Fonte: PNAD, microdados, IBGE.

Portanto os resultados sugerem que fatores associados à sindicalização foram responsáveis por uma diferença de 6 pontos percentuais em prejuízo da sindicalização num modelo de MQO pressuposto o propensity score.

## Conclusão

Este trabalho teve como preocupação central explicitar as dificuldades e problemas para a produção de dados empíricos nas ciências econômicas. Em geral, os pesquisadores de tal área se deparam com as circunstâncias já determinadas em que os fenômenos estudados se realizam, sendo assim, estudá-los é uma tarefa de caráter observacional, ou seja, uma tarefa que não pode se utilizar da reprodução experimental do fenômeno. Como decorrência de tal fato, o economista fica restrito do ponto de vista do instrumental analítico. Ao apresentar uma situação concreta permeada por tal problemática, o estudo do impacto da sindicalização sobre a renda do trabalhador, o presente trabalho pode discutir e aplicar o propensity score como uma das mais recentes soluções forjadas pela literatura de avaliação de tratamentos para minorar os erros de mensuração. O objetivo principal era reduzir o viés de seleção causado pelo fato de grupos de tratamento e controle poderem apresentar características sócio-econômicas significativamente distintas em estudos observacionais.

No primeiro capítulo foi fornecido um panorama geral do histórico dos sindicatos no país. As lutas que estes travaram por melhores condições de vida tiveram seus altos e baixos e se refletiram em diversos aspectos da vida social, entretanto muitos destes aspectos não se expressam sob a forma quantitativa. Por exemplo, a mudança para o regime democrático traz a retomada dos direitos civis, estas são conquistas qualitativas difíceis de serem quantificadas e não somente um fruto da ação sindical. Entretanto, as lutas salariais são um dos principais eixos da atividade sindical e podem ter seu impacto mensurado quantitativamente, pois tal é a natureza em que se manifesta a renda do trabalho. Assim, pode-se utiliza-la como proxy dos efeitos da sindicalização.

A discussão sobre o planejamento de experimentos feita no capítulo dois permitiu evidenciar a importância de um bom desenho experimental, que seja adequado ao tipo de estudo que o pesquisador pretenda realizar, para que este não seja induzido a conclusões errôneas. Foram descritas as principais metodologias utilizadas em planejamentos de experimentos e as principais fontes de preocupação no que se refere à validação interna e externa dos experimentos. Por fim, foi indicado o *propensity score* como uma proposta de desenho quase-experimental adequada para minimizar o viés de seleção no estudo do impacto da sindicalização na renda do trabalho.

Posteriormente, no capítulo três, procurou-se demonstrar como estimativas para o impacto da sindicalização sobre o rendimento do trabalhador podem apresentar resultados distintos caso não sejam considerados possíveis vieses em estudos observacionais. Utilizando os microdados da PNAD 2006, a avaliação do impacto da sindicalização foi sendo construída através de comparação direta entre sindicalizados e não-sindicalizados, de estimativas de MQO e do *propensity score*.

Por fim, os resultados produzidos apontaram que os sindicalizados recebem um rendimento médio superior ao dos não sindicalizados em 68% para uma comparação direta entre os grupos, em 12% para as estimativas de MQO e em 18% para o *propensity score*. É importante ressaltar que a análise feita no presente trabalho não considerou o efeito difusor dos ganhos da luta dos sindicalizados para os não sindicalizados. As divergências dos resultados demonstram como é importante conhecer o instrumental analítico, não só para corroborar hipóteses elaboradas teoricamente, mas também evitar fontes de vieses que podem surgir no uso dessas ferramentas.

## Referências

- ALVES, G. (1996). Nova ofensiva do capital, crise do sindicalismo e as perspectivas do sindicalismo no Brasil – o Brasil nos anos noventa. In: TEIXEIRA, F. & OLIVEIRA, M. **Neoliberalismo e reestruturação produtiva**. São Paulo: Cortez/UECE.
- BOITO JR., A. (1996). **Hegemonia neoliberal e sindicalismo no Brasil**. São Paulo: Crítica marxista, v. 1, n. 3, p. 80-105.
- CAMERON, A. C.; TRIVEDI, P. K. (2005). **Microeconometrics: Methods and Applications** New York: Cambridge University Press.
- CAMPBELL, D. T.; STANLEY, J. C. (1966). **Experimental and quasi-experimental designs for research**. Boston: Houghton Mifflin Company.
- FUKUYAMA, F. (1992); **Fim da historia e o último homem**. Rio de Janeiro : Rocco.
- HALVORSEN, R.; PALMQUIST, R. The interpretation of dummy variables in semilogarithmic equations. **The American Economic Review**, v. 70, n. 3, p. 474-475, jun. 1980.
- HECKMAN, J. J.; ICHIMURA, H.; Todd, P. E. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. **Review of Economic Studies**, vol. 64(4), pp. 605-54, October.
- HOFFMANN, R. (1998). Equação de rendimento para pessoas ocupadas no Brasil: contrastes regionais e setoriais. In: **Anais do XXXVI Encontro da SOBER**. Poços de Caldas/MG, p.707-717.
- HOFFMANN, R. (2000). Mensuração da desigualdade e da pobreza no Brasil. In: HENRIQUES, R. (Org). **Desigualdade e Pobreza no Brasil**. Rio de Janeiro: IPEA, p. 81-107.
- MARTINHO, F. P.; SANTANA, M. A. (2003). **Sindicato e processos de redemocratização no Brasil: analisando algumas conjunturas**. Lisboa: Penélope. Oeiras - Portugal, v. 27, p. 75-90.
- MATTOS, M. B (2003). **O sindicalismo brasileiro após 1930**. Rio de Janeiro: Jorge Zahar.
- NEDER, H. D.; SILVEIRA, J.M.F.J. da; BUAINAIN, A. M.; SOUZA FILHO, H. M.; MAGALHÃES, M.; PIRES, D.S.( 2002). Avaliação do Processo de Seleção e Efeitos Imediatos dos Programas de Reforma Agrária: Cédula da Terra – PCT e Assentamentos do INCRA – In: **Anais do XXX Encontro Nacional da ANPEC**. CD ROM.
- RAVALLION, M.; (1999). **The mystery of the vanishing benefits: Ms. Speedy Analyst's introduction to evaluation**. Washington: World Bank.
- RESENDE, A. C. C.; OLIVEIRA, A. M. H. C. de. (2008). **Avaliando resultados de um programa de transferência de renda: o impacto do Bolsa-Escola sobre os gastos das famílias brasileiras**. Estudos Econômicos, v. 38, n. 2, pp. 235-265.

ROSENBAUM, P. R.; RUBIN, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. **Biometrika**, v. 70, pp. 41–55.

SILVA, J. E. (2002). **O Nascimento da organização sindical no Brasil e as primeiras lutas operárias (1890-1935)**. Disponível em: <<http://www.cecaiorg.br/edgar/PrimeirasLutas.htm>>. Acessado em Acesso em: <22 de agosto de 2002>