

# AS CONDIÇÕES DE SAÚDE AFETAM OS RENDIMENTOS DO TRABALHO? EVIDÊNCIAS PARA O MERCADO DE TRABALHO NO BRASIL

## Área 13 - Economia do Trabalho

**Wallace Patrick Santos de Farias Souza**

PPGE – UFRGS. E-mail: wpsfarias@gmail.com

**Flávio Augusto Ziegelmann**

PPGE – UFRGS. Pesquisador do CNPq. E-mail: flavioz@ufrgs.br

**Erik Figueiredo**

PPGE – UFPB. Pesquisador do CNPq. E-mail: eafigueiredo@gmail.com

**Resumo:** O objetivo deste estudo consiste em mensurar o efeito das condições de saúde sobre o rendimento dos trabalhadores no mercado de trabalho brasileiro, usando dados da PNAD de 2008, que traz um suplemento de saúde individual. Para tanto, adotou-se uma estratégia empírica baseada em quatro passos: i) estimar o efeito de tratamento médio por métodos de pareamento com correção semiparamétrica para o viés de seleção amostral; ii) efeito de tratamento quantílico com correção do viés de seleção; iii) método de Lewbel (2012) que gera instrumentos internos a partir da heterocedasticidade dos resíduos e iv) *bounds* para o efeito do tratamento seguindo Oster (2015). Foram utilizados como tratamento quatro especificações de condições de saúde. Os resultados mostram um impacto negativo de uma má condição de saúde na renda do trabalho para todas as especificações utilizadas, sendo que os indivíduos que sofrem as maiores reduções salariais são justamente os que estão nos quantis mais baixos de renda. Isso agrava mais ainda sua condição criando um “círculo vicioso”. Tais resultados são corroborados pelo método de Lewbel (2012) usado quando não se tem instrumentos externos disponíveis. Por fim, a estimação dos *bounds* fornece robustez aos resultados, e o coeficiente de proporcionalidade mostra que o viés causado pelas não observáveis não é suficiente para invalidar os resultados.

**Palavras-Chave:** Condição de Saúde, Efeito de Tratamento, Renda do Trabalho.

**Abstract:** The objective of this study is to measure the impact of health conditions on the income of workers in the Brazilian labor market, using data from PNAD 2008, which has an individual health supplement. Therefore, we adopted an empirical strategy based in four steps: i) estimate the average treatment effect by matching methods with semi-parametric correction for sample selection bias; ii) quantile treatment effect with correction of selection bias; iii) Lewbel (2012) method that generates domestic instruments from residuals heterocedasticity and iv) bounds for the effect of the treatment following Oster (2015). We used as treatment four specifications for health conditions. The results show a negative impact of a bad health condition in labor income for all specifications used, and individuals suffering the biggest pay cuts are precisely those in the lowest income quantile. This aggravates their condition even more by creating a "vicious circle". These results are corroborated by Lewbel method (2012), used when there is no external instruments available. Finally, the estimation of bounds provides robustness to results and the proportionality coefficient shows that the bias caused by unobservable variables is not sufficient to invalidate the results.

**Keywords:** Health Condition, Treatment Effect, Labor Income.

**JEL Classification:** C21, I10, J22.

## 1. Introdução

O acesso à saúde é considerado um dos principais fatores de acumulação de capital humano e influencia positivamente na função de bem-estar dos indivíduos, dado que quanto maior seu estoque de saúde maior a disponibilidade para o trabalho, estudo e/ou lazer. A condição saudável é também considerada um fim em si mesmo, pois só o fato de ter uma boa saúde contribui para o aumento do bem-estar dos indivíduos. As condições de saúde afetam os rendimentos dos trabalhadores na medida em que influenciam a sua decisão em participar ou não da força de trabalho, o número de horas trabalhadas e a sua produtividade. Dessa forma, indivíduos saudáveis possuem um maior estoque de capital humano e tendem a ser mais produtivos quando comparados aos seus pares não saudáveis. (ALVES, 2002).

A saúde, diferente de outras formas de capital humano, tem uma importância peculiar para o retorno econômico, pois as condições físicas e mentais com que o indivíduo se encontra determinam a quantidade e intensidade com que os mesmos podem se dedicar a atividade laboral, contribuindo para o aumento dos rendimentos e bem-estar (GROSSMAN, 1972). Trabalhos pioneiros como Schultz (1961) e Mincer (1974) apontam saúde e educação como componentes principais do capital humano.

Fatores não-controláveis, como herança genética, e controláveis, como hábitos saudáveis, definem o estoque de saúde dos indivíduos, e este determina o quanto de trabalho os indivíduos estão dispostos e são capazes de ofertar. O tipo de atividade é também importante para determinar a quantidade de trabalho ofertada, pois a presença de qualquer doença que gere limitações físicas poderá ter um impacto maior em

tarefas que exijam um maior esforço físico, que por sua vez, são normalmente tarefas desempenhadas com menor grau de qualificação (Noronha e Andrade, 2004).

Grossman (1972) introduz a ideia de que a procura de saúde é resultado de um processo de escolha individual. Assim, os cuidados médicos são entendidos como um fator produtivo adquirido pelos indivíduos para produzir saúde. Para o autor, a saúde é um estoque, tendo duração de vários anos, e tal estoque se depreciaria com o passar do tempo. O modelo proposto pelo autor evidencia que a procura de cuidados médicos terá um objetivo último que é a procura de saúde e, além disso, a procura por cuidados médicos é influenciada por fatores sócio-econômicos que afetam o estado de saúde, tais como, as preferências individuais, o salário, a idade, o nível educacional.

A grande maioria dos trabalhos que analisam os determinantes do retorno salarial do indivíduo focam na educação como componente de capital humano, tal como na formulação de Mincer (1974). Outros trabalhos como Bartel & Taubman (1979), Levine et al. (1997), Kassouf (1997), Alves e Andrade (2003) e Bockerman et al. (2014), acrescentam um componente de saúde na formulação minceriana, sobre a hipótese que o estoque de saúde afeta a produtividade do trabalho e oferta de mão de obra.

Para o Brasil, apesar de ser uma área recente de pesquisa, alguns trabalhos se destacam como referência na análise do impacto da saúde no resultado econômico individual. Kassouf (1999), por exemplo, verifica que um IMC<sup>1</sup> baixo (usado como indicador de subnutrição) tem um impacto significativo sobre o rendimento dos indivíduos para o Brasil, enquanto que Alves e Andrade (2003), em estudo para o Estado de Minas Gerais, avalia que as más condições de saúde tem como efeitos a redução salarial ou até a saída do indivíduo do mercado de trabalho.

A maioria dos artigos identificam três canais pelos quais as condições de saúde afetam as decisões dos indivíduos no mercado de trabalho e consequentemente os ganhos auferidos por tais trabalhadores: a participação na força de trabalho, o número de horas trabalhadas e o nível de salário. Com respeito à oferta de trabalho, a saúde tem impactos sobre a escolha setorial, sobre a decisão de trabalhar e sobre número de horas a ser ofertado (STRAUSS & THOMAS, 1998). No entanto, o resultado de uma má condição de saúde sobre a oferta de trabalho pode ser ambíguo, dependendo de três efeitos, a saber: efeito dotação<sup>2</sup>, efeito substituição<sup>3</sup> e efeito renda<sup>4</sup>. Se, por exemplo, o efeito renda superar o efeito dotação mais o efeito substituição juntos, uma saúde precária resulta em maior oferta de trabalho. Mas no geral, é esperado que aconteça o contrário, e que um trabalhador doente tende a trabalhar menos tempo do que uma pessoa saudável.<sup>5</sup>

Por outro lado, indivíduos que possuem doenças (sobretudo crônicas) buscam se adaptar ao ambiente de trabalho e executar normalmente suas tarefas, a não ser que a doença seja grave o suficiente que o impeça de trabalhar. Para saber se de fato o estado de saúde implica em restrições capazes de reduzir a produtividade e perdas salariais, deve ser levada em conta a gravidade da enfermidade e a capacidade de adaptação do indivíduo.

Com relação à oferta de trabalho, esta é dada pelo número de horas trabalhadas por semana, que corresponde à soma das horas semanais alocadas para o trabalho principal, secundário e outros. Porém, esse número de horas assim como o nível de rendimento é observado apenas para aqueles que participam da força de trabalho, de maneira que a amostra utilizada na estimação das equações não é aleatória, dado que indivíduos que não participam da força de trabalho não aparecem na amostra selecionada. Dessa forma, é preciso um método que corrige a seletividade amostral, observando quais variáveis afetam a probabilidade de um determinado indivíduo entrar no mercado de trabalho. Muitos trabalhos aplicados tem usado a correção do viés de seleção baseado no procedimento de Heckman (1979), segundo o qual as motivações que os indivíduos têm para ocupar postos de trabalho podem diferir entre os indivíduos, levando em conta suas características.

Em suma, as condições de saúde podem afetar a participação no mercado de trabalho, o nível de salário por hora individual via efeitos sobre a produtividade e a quantidade de horas trabalhadas pelos

---

<sup>1</sup> Índice de Massa Corporal.

<sup>2</sup> Refere-se a menor quantidade de tempo para realizar as atividades, inclusive trabalhar, quando esta doente.

<sup>3</sup> Quando a condição de saúde precária corresponde a uma taxa de salário menor, fazendo com que o indivíduo substitua trabalho por lazer (reduza sua oferta de trabalho), dado que o trabalho gera desutilidade.

<sup>4</sup> Quando a condição de saúde precária faz o indivíduo trabalhar mais para compensar a menor taxa de salário por hora.

<sup>5</sup> A intensidade com que essa redução acontece pode depender, entre outros fatores, do tipo de doença.

indivíduos. Para o presente trabalho, o foco será no efeito que as condições de saúde exercem sobre os rendimentos do trabalho, sendo a variável dependente o salário por hora semanal auferido, em logaritmo.

Dito isso, o objetivo deste artigo é mensurar o impacto das condições de saúde dos indivíduos no seu nível salarial, a partir dos microdados da Pesquisa Nacional por Amostra de Domicílios para o ano de 2008, que contem um suplemento de saúde dos entrevistados<sup>6</sup>. Para alcançar tal objetivo, as estimativas serão baseadas em quatro estratégias empíricas, como forma de tentar mitigar a possibilidade de variáveis não observadas estarem viesando os resultados. A primeira delas é estimar o efeito de tratamento médio das condições de saúde sobre a renda do trabalho através do *propensity score matching*, como forma de ter estimativas mais razoáveis que as obtidas por OLS.<sup>7</sup> Em seguida será estimado o efeito por quantis, como forma de ver o impacto do tratamento (*status* de saúde) por toda a distribuição de salários. Para os dois casos é feita uma estimação em primeiro estágio para corrigir o viés de seleção, verificando as variáveis que afetam a probabilidade do indivíduo estar inserido no mercado de trabalho. Dado a ausência de instrumentos externos disponíveis, em terceiro lugar é usado um estimador de dois estágios proposto por Lewbel (2012) que explora a heterocedasticidade do primeiro estágio para gerar instrumentos para identificação; e por fim, serão estimados os *bounds* para o efeito de tratamento desenvolvido por Oster (2015) que informa sobre o viés que persiste devido a variáveis não observadas, dada a preocupação com variáveis omitidas na estimação do coeficiente de interesse.

O artigo apresenta a seguinte estrutura, além desta introdução. A seção seguinte traz um resumo das evidências empíricas acerca da influência das condições de saúde no mercado de trabalho. Na terceira seção é feita uma análise do estado de saúde no Brasil, segundo a amostra utilizada, a descrição da base de dados e a criação das variáveis de saúde usadas como tratamento. A quarta seção apresenta todos os passos da estratégia empírica adotada, bem como o método de correção do viés de seleção. Em seguida são apresentados os resultados encontrados e, por fim, na sexta seção são feitas as considerações finais.

## 2. Revisão de Literatura

Vários trabalhos empíricos têm sido realizados para tentar entender a relação das condições de saúde e mercado de trabalho, seja usando a saúde como uma variável exógena tal qual em Luft (1975) e Kassouf (1999), ou tentando controlar a endogeneidade entre saúde e rendimentos tal como Grossman e Benham (1974) e Schultz e Tansel (1997). O trabalho de Luft (1975) mensura as perdas de salários ocasionadas por uma má condição de saúde dos trabalhadores americanos entre 18 e 64 anos, diferenciando por sexo e raça. Kassouf (1999) utiliza o mesmo procedimento para o Brasil mensurando as condições de saúde através do estado nutricional individual, avaliando também as diferenças regionais e entre as áreas urbana e rural.

Segundo Gomes et al (2012), normalmente indivíduos com menores rendimentos desempenham atividades que exigem maior esforço físico que intelectual, o que fundamenta a relação dos rendimentos condicionados ao estado de saúde. Dessa forma, os autores identificam um “círculo vicioso” em tal relação, onde os indivíduos mais pobres tendem a ocupar trabalhos que requerem maior esforço e conseqüentemente maior perda de rendimentos quando se encontram doentes (devido à impossibilidade de empreender o esforço), o que pode torná-los mais pobres e agravar mais ainda a condição de saúde. Assim, a presença de qualquer tipo de doença que gere limitações físicas poderá ter um impacto maior sobre os rendimentos dos indivíduos que desempenham tarefas que demandem mais esforço físico, quando comparadas a atividades que requeiram um menor nível de esforço (Noronha e Andrade, 2004). Em outras palavras, visto que a remuneração pode ser definida pela produtividade, indivíduos com condições de saúde precária tendem a ser menos produtivos. Isso implica que irão auferir uma menor remuneração, a depender do tipo de doença.

Nesse sentido, usando equações simultâneas para dados da Austrália da *AusDiab Survey* (ADS), Cai (2010) verificou que os dois efeitos podem ser encontrados. A saúde afeta a participação da força de trabalho e por outro lado a participação da força de trabalho afeta a saúde, sendo tal efeito negativo para os homens e positivo para as mulheres. Em âmbito internacional, alguns trabalhos tem usado painel de dados quando a disponibilidade dos mesmos torna possível, tais como Disney et al (2006) e García-Gómez (2010). Disney et al (2006) estimam um painel de dados da *British Household Panel Survey* (BHPS) de 1991-1998

---

<sup>6</sup> Os detalhes sobre a Base de dados utilizada será explicitado na seção 3.

<sup>7</sup> Ordinary least squares

para o Reino Unido e encontra que choques negativos na condições de saúde tem impacto sobre a aposentadoria precoce dos trabalhadores e consequentemente um menor tempo no mercado de trabalho. enquanto que García-Gomez et al (2010) usando dados da mesma pesquisa para 1991-2002 mostram que a saúde afeta a entrada e saída do emprego.

Por outro lado, Morris (2007) aplica vários métodos (*propensity score matching*, *probit*, *probit* bivariado com variável instrumental) para dados da *Health Survey for England* de 1997-1998 e verifica um impacto negativo da obesidade sobre o nível de emprego, enfocando que a estimativa que não corrige a endogeneidade com o uso do instrumento é subestimada para as mulheres.

Para o Brasil, Alves e Andrade (2003) encontram efeitos diferenciados da saúde nos rendimentos de homens e mulheres no mercado de trabalho para Minas Gerais. Para os homens o principal efeito de uma má condição de saúde é a exclusão do mercado de trabalho, enquanto que para as mulheres é a redução na taxa de salários. Seguindo essa linha, Godoy et al (2006) encontram efeitos negativos na renda individual devido a doença renal crônica a partir dos dados da PNAD de 1998, impactando com mais intensidade trabalhadores de baixa renda. A hipótese dos autores é que além de impactos no mercado de trabalho propriamente dito<sup>8</sup>, a doença real tem outras implicações econômicas como aposentadoria precoce, aumento de programas assistenciais de transferência de renda e custos elevados do sistema de saúde. Cuidados com prevenção e o retardamento dessa e outras doenças estão entre as recomendações para a redução das disparidades de rendimentos no mercado de trabalho. Silva (2011) encontra resultados semelhantes com relação à tuberculose, que implica em perdas salariais para os trabalhadores, sobretudo os mais pobres.

Quando considerados no agregado, tais efeitos sobre os rendimentos individuais podem influenciar as condições macroeconômicas. Posto isso, Figueiredo et al (2003) verificaram durante a década de 1990 que um bom estado de saúde da população contribui positivamente para o crescimento econômico, seja através da taxa de crescimento do PIB ou reduções na taxa de mortalidade. Os autores argumentam que maiores investimentos em capital humano podem ampliar tais efeitos. Em suma, a literatura tem corroborado que uma má condição de saúde afeta negativamente a participação no mercado de trabalho, seja pela exclusão do trabalhador ou redução do rendimento e jornada de trabalho.

### 3. Base de dados

#### 3.1 Análise do Estado de saúde no Brasil

A Pesquisa Nacional por amostra de Domicílios de 2008 trás um suplemento com informações sobre o estado de saúde dos indivíduos, que pode ser mensurado a partir de vários aspectos e critérios. Seguindo a literatura e a disponibilidade de dados, a avaliação empírica do estado de saúde pode ser autopercebida, baseada em conhecimentos e crenças pessoais, ou reportada por um profissional de saúde, observando a incidência de uma série de doenças crônicas que os indivíduos possam apresentar. São também utilizadas variáveis que indicam se o indivíduo apresenta problemas de realizar tarefas ou se locomover e ainda aspectos comportamentais relacionados à saúde, como tabagismo e sedentarismo/atividade física.

Com relação à autoavaliação reportada pelo indivíduo, este é um critério subjetivo que, no entanto capta a percepção do indivíduo sobre sua real disposição em realizar determinadas tarefas, além de ser um indicador sobre o estado de saúde global. Uma desvantagem é que pode está relacionado com características socioeconômicas e culturais das pessoas.

Já os indicadores de dificuldades de mobilidade ou limitações funcionais, apesar de também serem autoreportados, estão menos sujeitos a subjetividade dado que focam em perguntas específicas sobre determinadas tarefas corriqueiras, que o indivíduo tem consciência sobre o seu real desempenho. Um problema é que tais dificuldades apresentam baixa frequência, sobretudo na população mais jovem, o que torna difícil captar o estado de saúde nessa faixa etária.

A Tabela 1 apresenta o estado de saúde autoavaliado, onde o indivíduo é questionado como em geral classifica seu próprio estado de saúde, sendo possíveis cinco respostas (muito bom, bom, regular, ruim, muito ruim). É percebido que enquanto 3% dos entrevistados classificam o seu estado de saúde ruim ou muito ruim, mais de 70% o classificam como muito bom ou bom, chegando a quase 80% para os homens.<sup>9</sup>

<sup>8</sup> Tais como redução de salários e de número de horas trabalhadas.

<sup>9</sup> Para fins metodológicos, como será detalhado mais adiante, tal variável será transformada em binária, que assume valor 0 para os indivíduos que apresentam o estado de saúde autodeclarado como “muito bom” ou “bom”, e 1 para indivíduos que apresentam o estado “regular”, “ruim” ou “muito ruim”.

O alto percentual de pessoas que consideram sua condição de saúde como boa (muito boa) pode indicar o desconhecimento em relação a alguma doença que não foi diagnosticada ou a própria reluta em admitir alguma enfermidade, sobretudo para os homens, menos propensos a realizar consultas de rotina e cuidados gerais com a sua saúde.

**Tabela 1: Autoavaliação do Estado de Saúde Individual – PNAD (2008) – em %**

Autoavaliação de Saúde	Homens	Mulheres	Total
Muito Bom	23,07	19,66	21,35
Bom	54,85	53,79	54,32
Regular	19,51	23,33	21,44
Ruim	2,23	2,60	2,42
Muito Ruim	0,34	0,62	0,48

Fonte: Elaboração dos autores com base nos dados da PNAD (2008)

Outro critério a ser utilizado é o clínico/funcional avaliado por um profissional de saúde, um indicador importante para avaliar o estado de saúde individual pelo fato de fornecer uma informação objetiva para mensurar a saúde. No entanto, se o indivíduo não tem costume de ir ao médico ou fazer exames periódicos, pode ser que desconheça a existência de algum problema de saúde que possa ter, principalmente aqueles que não apresentam sintomas claros em fases iniciais.

Entre as doenças avaliadas, são observadas as de características como alta prevalência (hipertensão), altas taxas de utilização de serviços de saúde (bronquite e asma), possibilidade de intervenção efetiva (tuberculose) e elevada frequência de uso de serviços de alto custo (insuficiência renal crônica, câncer). Ao todo são 12 doenças para as quais os indivíduos são ou não diagnosticados tê-las por um profissional de saúde, sendo o percentual de indivíduos identificados com cada uma dessas doenças reportado na Tabela 2. É importante salientar que os indivíduos podem apresentar mais de uma doença, fato não destacado na Tabela, mas que será captado pela criação de uma variável referente ao número de doenças<sup>10</sup>.

**Tabela 2: Indivíduos diagnosticados com alguma doença – PNAD (2008) – em %**

Doenças	Homens (% do total)	Mulheres (% do total)	Total
Coluna	43,62	56,38	19,27
Artrite	33,61	66,39	6,22
Câncer	37,10	62,90	0,46
Diabetes	48,42	51,58	3,26
Bronquite	35,86	64,14	3,52
Hipertensão	43,42	56,58	15,76
Coração	44,22	55,78	3,44
Insuficiência Renal	44,23	55,77	1,52
Depressão	23,41	76,59	5,27
Tuberculose	53,57	46,43	0,14
Tendinite	24,11	75,89	4,61
Cirrose	71,23	28,77	0,12

Fonte: Elaboração dos autores com base nos dados da PNAD (2008)

Pela Tabela 2 é possível perceber que doenças relacionadas à coluna e Hipertensão são as mais frequentes na população brasileira, enquanto que cirrose e tuberculose afetam uma parcela muito pequena. Quanto à diferença de gênero, as mulheres representam a maioria do total de afetados para praticamente todas as doenças, chegando a 76% do total de pessoas com depressão. Apenas a tuberculose e a cirrose

<sup>10</sup> A criação dessa variável será explicada mais adiante.

afetam uma proporção maior de homens, esta última chegando a mais de 71% do total de entrevistados com a doença.

**Tabela 3: Autoavaliação do Estado de Saúde de Indivíduos diagnosticados com alguma doença – PNAD (2008) – em %**

Doenças	Muito Bom	Bom	Regular	Ruim	Muito Ruim
Coluna	9,57	42,38	40,18	6,49	1,38
Artrite	5,63	33,67	47,71	10,65	2,34
Câncer	9,54	34,28	43,11	12,01	1,06
Diabetes	4,77	31,76	51,08	9,48	2,91
Bronquite	12,63	43,99	35,90	5,99	1,49
Hipertensão	6,64	41,93	43,17	6,75	1,52
Coração	4,76	27,10	51,93	12,93	3,28
Insuficiência Renal	5,72	31,07	49,08	11,22	2,91
Depressão	7,06	34,82	44,64	10,91	2,58
Tuberculose	13,10	42,86	35,71	8,33	0,00
Tendinite	13,85	43,47	34,94	6,14	1,60
Cirroose	9,59	38,36	31,51	17,81	2,74

Fonte: Elaboração dos autores com base nos dados da PNAD (2008)

Uma mesma doença pode afetar as pessoas com diferente intensidade e causar diferentes percepções quanto ao estado de saúde individual. Dessa forma, a Tabela 3 indica como os indivíduos diagnosticados com alguma doença classificam seu estado de saúde. Para todas as doenças, observa-se que a maioria dos indivíduos diagnosticados reporta seu estado de saúde como bom ou regular o que pode indicar a perda de rendimentos ou redução do número de horas trabalhadas, mas não a exclusão do mercado de trabalho. Esse fato é corroborado pela estatística descritiva mais adiante, onde a quase totalidade de indivíduos da amostra esta inserida no mercado de trabalho.

As dificuldades de mobilidade e/ou locomoção, a criação das variáveis binárias que serão usadas como tratamento e as demais características individuais e de localidade usadas como controle serão descritas junto com as principais estatísticas descritivas abaixo.

### 3.2 Base de Dados:

A base de dados utilizada neste estudo são os microdados da Pesquisa Nacional por Amostra de Domicílios (PNAD) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para o ano de 2008. Além das informações permanentes sobre características socioeconômicas domiciliares e pessoais (como renda, idade, raça, sexo, renda per capita, etc.), esse ano investiga características do levantamento suplementar de saúde. O suplemento é uma fonte de dados importante para o conhecimento e monitoramento de aspectos relevantes da situação de saúde da população brasileira e do acesso, utilização e financiamento de serviços de saúde, assim como informações sobre mobilidade, sedentarismo, entre outros.

Foram utilizados dois critérios para classificar os indivíduos entre doentes ou saudáveis, além de variáveis usadas como controle que indicam a existência de dificuldade de problemas de locomoção e execução de tarefas domésticas e corriqueiras. Tais critérios foram recodificados para variáveis binárias e usados como tratamento para diferenciar os indivíduos quando ao impacto na renda do trabalho e participação no mercado de trabalho.

O primeiro critério é a condição de saúde autoavaliada, que embora seja uma avaliação subjetiva dos próprios indivíduos, dá uma noção de percepção global do estado de saúde e é consistente a diferentes medidas autoreportadas. As respostas variam de 1 a 5 de acordo com a seguinte classificação: 1= “muito boa”, 2= “boa”, 3= “regular”, 4= “ruim”, 5= “muito ruim”. Dessa forma, ao contrário de Noronha e Andrade (2007), foi criada uma variável *dummy* que assume valor um se indivíduo relata sua condição de

saúde como “regular”, “ruim” ou “muito ruim”, e assume valor zero para uma condição “boa” ou “muito boa”, fazendo com que o grupo de controle seja saudável<sup>11</sup>.

O outro critério de saúde adotado como tratamento é a presença de doença diagnosticada por um profissional de saúde, sendo uma variável *dummy* que assume valor um se o indivíduo foi diagnosticado com uma ou mais doenças e zero caso não tenha nenhuma doença. Como robustez, foram construídas *dummies* (também usadas como tratamento em outras estimativas, uma por vez) para verificar se uma pessoa possui mais de uma doença, ou seja, uma variável binária assumindo valor um se possui duas ou mais doenças e zero caso saudável, e por fim, uma variável binária que assume valor um se o indivíduo possui três ou mais doenças e zero caso não tenha sido diagnosticado com nenhuma doença.

O critério subjetivo como um indicador da percepção global que o indivíduo tem a respeito de seu estado de saúde, possivelmente, mensura de forma mais adequada o estado de saúde do indivíduo do que o critério clínico/ funcional, e estaria captando de forma mais ampla as restrições que o estado de saúde impõe ao desempenho dos doentes no trabalho, sobretudo, sobre a sua produtividade.

Com relação à restrição de atividades rotineiras por motivo de saúde, refere-se a comportamentos geralmente associados à redução de atividades que a pessoa é capaz de desempenhar no seu dia a dia. Tais restrições<sup>12</sup> foram transformadas em variáveis binárias que assumem o valor um se o indivíduo apresenta dificuldade para aquela característica e zero caso contrário, sendo depois agrupadas em uma variável que representa o número de problemas/dificuldades de realizar atividades e/ou locomoção que o indivíduo enfrenta. Tal variável foi usada como controle em todas as estimações.

Características pessoais (sexo, raça, idade, escolaridade, hábitos saudáveis), relacionadas à região (UF, área metropolitana, etc) e informações sobre o setor e ramo de atividade no mercado de trabalho foram também usadas como controles. Por fim, para a equação de participação no mercado de trabalho que corrige o viés de seleção, além das características pessoais e de região citadas foram utilizadas variáveis de renda do não trabalho (pensão, juros, aposentadorias, aluguel, etc), que são incluídas por afetarem o salário de reserva dos agentes e a decisão de entrar ou não no mercado de trabalho,<sup>13</sup> ou seja, os indivíduos exigem salários de reserva maiores para trocar lazer por trabalho. Estas são as variáveis de exclusão do modelo, dado que afetam a probabilidade de um indivíduo ofertar ou não trabalho, mas não afetam diretamente os rendimentos<sup>14</sup>.

Além das variáveis relacionadas às condições de saúde já apresentadas, a Tabela A.1 em anexo apresenta as principais estatísticas descritivas da amostra utilizada, num total de 60.812 observações sobre indivíduos da PNAD de 2008. É observado que a quase totalidade dos indivíduos está inserido de alguma forma no mercado de trabalho, sendo esta a variável dependente da equação de participação (1º estágio) que corrige o viés de seleção amostral. Para a equação de interesse (2º estágio) a renda do trabalho usada para construir o logaritmo do salário por hora apresenta uma média de R\$ 1.211,93. Na tabela A.1 são marcadas as variáveis usadas em cada um dos estágios.

Os diferentes tipos de renda do não trabalho apresentam uma média muito baixa e um grande desvio padrão, devido ao fato que muitos indivíduos não recebem tais tipos de rendimentos. Com relação aos atributos pessoais, há um maior percentual de pessoas não brancas e uma leve participação a mais de mulheres na amostra, com uma média de idade e escolaridade de 40 anos e 9,23 anos de estudo, respectivamente. Foram excluídas pessoas com menos de 15 e mais de 65 anos de idade, com o intuito de captar apenas os indivíduos com idade ativa para o trabalho, o que faz a média de idade e escolaridade ser um pouco elevada em comparação com a amostra completa. Ainda é notado que do total da amostra, cerca de 28% praticam alguma atividade física e 20% são fumantes, sendo hábitos que podem influenciar no *status* de saúde.

<sup>11</sup> A definição como foi exposto é para padronizar com a outra variável binária de condição de saúde, que assume valor 1 se o indivíduo é diagnosticado com alguma doença.

<sup>12</sup> As características de locomoção são relacionadas a se o indivíduo tem: dificuldade em alimentar-se, dificuldade de levantar objetos, dificuldade em realizar consertos domésticos, dificuldade de subir ladeira ou escada, dificuldade de abaixar-se, dificuldade de andar mais de 100 metros, dificuldade para andar mais de um quilômetro.

<sup>13</sup> Quanto maior a renda do não trabalho, maior tenderá a ser o salário de reserva das pessoas, que consequentemente esperam um salário maior para aceitar uma ocupação.

<sup>14</sup> Ao contrário do método de correção proposto por Heckman (1979), será utilizado um método semiparamétrico onde tal restrição não é obrigatoriamente necessária. Mesmo assim, foram utilizadas as rendas do não trabalho, dada a justificativa teórica.

Dentre os aspectos relacionados ao mercado de trabalho, a maioria encontra-se no comércio e na indústria de transformação, totalizando 30% do total de trabalhadores. Com relação ao *status* da posição de ocupação, os trabalhadores com carteira assinada e por conta própria possuem uma representação em torno de 31% e 27%, respectivamente. Por fim, como é de se esperar, a grande maioria dos indivíduos reside na área urbana (86%) e nas regiões Sudeste e Nordeste do país.

#### 4. Estratégia Empírica

Esta seção apresenta a estratégia empírica adotada no estudo. Em primeiro lugar será descrito o método de efeito de tratamento médio por *propensity score matching* e o tratamento por quantis, além da correção do viés de seleção para o primeiro estágio. Em seguida, na subseção 4.2 será apresentado o método de Lewbel (2012), que usa instrumentos internos a partir da heterocedasticidade, quando não existem restrições de exclusão disponíveis para gerar as estimativas. Por fim, a subseção 4.3 apresenta o método de Oster (2013, 2015) para estimar *bounds* para o efeito de tratamento, como robustez para as estimações anteriores.

##### 4.1 Efeito de Tratamento

O interesse do presente trabalho é estimar o efeito causal das condições de saúde sobre as perdas de rendimento do trabalho no Brasil, para indivíduos que reportam seu estado de saúde como ruim ou que tem alguma doença diagnosticada por um profissional de saúde (efeito nos tratados), usando dados do suplemento de saúde da PNAD, como destacado anteriormente. Para tanto, como especificação tradicional para estimar tal efeito, tem-se

$$Y = \alpha + \beta S + X\gamma + \varepsilon, \quad (1)$$

onde  $Y$  é o log da renda do trabalho (salário) por hora,  $S$  é um indicador (*dummy*) para saúde,<sup>15</sup>  $X$  é um vetor de variáveis de controle que afetam os salários, e  $\varepsilon$  é um termo de erro. Mensurar  $\beta$  de forma consistente requer que o termo de erro não esteja correlacionado com a variável de saúde ( $\text{COV}(S; \varepsilon) = 0$ ), ou seja, que os indivíduos sejam distribuídos aleatoriamente pelas condições de saúde (condicionado as variáveis observadas), e não satisfazer tal condição torna as estimativas por OLS viesadas e inconsistentes (Angrist e Pischke, 2009).

Quando a hipótese não se mantém e a seleção do tratamento é baseada em variáveis não observadas correlacionadas com a variável de interesse ( $S$ ), uma das soluções seria encontrar uma variável instrumental para o efeito causal da condição de saúde que não seja correlacionada com o resultado. No entanto, como nem sempre encontrar um instrumento é trivial, é necessária uma estratégia de identificação que consiga estimar o efeito de interesse da condição de saúde sobre o rendimento do trabalho.

A primeira estratégia adotada, comumente usada para balancear variáveis em busca de resultados mais precisos, é o estimador por *propensity score matching*. Dado o conjunto de variáveis disponíveis, estimar o escore de propensão associado à condição de saúde reduz a influência de variáveis potenciais omitidas. (Angrist e Haid, 2004).

Segundo Rosenbaum e Rubin (1983) denota-se  $Y_i$  como o resultado observado do indivíduo  $i$  para a variável de resultado, o log do salário por hora, e  $Y_1$  e  $Y_0$  são os resultados potenciais para o grupo de tratamento (má condição de saúde ou diagnosticado com alguma doença) e que não possui a característica do tratamento (grupo de controle), respectivamente. Dessa forma, tem-se

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

De acordo com Imbens e Wooldridge (2009), sabe-se que não se podem comparar os dois resultados para um mesmo indivíduo num mesmo período de tempo, sendo observado apenas um dos dois resultados potenciais dado o status do tratamento,  $Y_i = Y_i^0 + (Y_i^1 - Y_i^0)S_i$ . Então, é preciso encontrar indivíduos

<sup>15</sup> Como mencionado na seção de dados serão utilizados vários indicadores de saúde como variável binária de tratamento, um por vez para cada estimação, a saber: i) *dummy* = 1 para má condição de saúde autoavaliada e 0 caso contrário; ii) *dummy* = 1 para indivíduo diagnosticado com pelo menos uma doença por um profissional de saúde e 0 caso contrário; iii) *dummy* = 1 para o indivíduo diagnosticado com duas ou mais doenças e 0 com nenhuma doença; iv) *dummy* = 1 para indivíduo diagnosticado com três ou mais doenças e 0 com nenhuma doença; e v) *dummies* para cada uma das doze doenças listadas nos dados, sendo 1 = tem a doença e 0 = não tem a doença.



pertencentes aos dois grupos que possam ser comparados (tratados e controle), depois de ajustadas às características observadas contidas em  $X_i$  (Angrist e Pischke, 2009).

Tal comparação é justamente a intuição dos estimadores de correspondência, onde, condicionando as variáveis  $X_i$ , os resultados potenciais são independentes do tratamento.<sup>16</sup> Ou seja,  $Y_i^1, Y_i^0 \perp\!\!\!\perp S_i | X_i$ , que implica em  $Y_i^1, Y_i^0 \perp\!\!\!\perp S_i | p(X_i)$ , onde  $p(X_i)$  é a probabilidade de um indivíduo ser tratado ou o *propensity score*. Isso permite encontrar o efeito de tratamento para cada valor de  $X_i$  como a diferença das médias dos dois *status* de resultado, usando para isto o *propensity score* para obter o efeito sobre os indivíduos tratados (Rosenbaum e Rubin, 1985). Assim, o ATT<sup>17</sup> pode ser obtido por

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

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

Para obter o  $\beta$  na equação (4) estima-se o escore de propensão utilizando um modelo probit, fazendo o *matching* do grupo de tratamento com o grupo de controle a partir da estimação por kernel (um indivíduo tratado comparado com uma ponderação de indivíduos do controle), e do vizinho mais próximo (um indivíduo tratado comparado com um indivíduo do controle). Todas as estimações serão feitas em dois estágios, onde o primeiro estágio é usado para correção do viés de seleção, observando variáveis que alteram a probabilidade do indivíduo está ou não inserido no mercado de trabalho.<sup>18</sup>

No entanto, o efeito médio pode não refletir completamente a influência do tratamento sobre os resultados potenciais, visto que a distribuição da variável dependente pode ser diferente na parte superior da inferior da distribuição, dependendo das variáveis utilizadas (o que faz sentido quando se trata de rendimentos). Dito isso, será estimado também o efeito do tratamento quantílico (QTEs),<sup>19</sup> dado seu poder intuitivo e a vantagem em permitir caracterizar o efeito em toda a distribuição condicional da variável resposta a partir de um conjunto de regressores. Além dessa vantagem, de acordo com Koenker (2005) apud Silva (2011) a utilização da regressão quantílica também pode ser usada quando a distribuição não é gaussiana; a regressão é robusta a outliers por utilizar a distribuição condicional da variável resposta; os estimadores podem ser mais eficientes que os obtidos por meio de OLS, dado que os erros não possuem uma distribuição normal; entre outras. É importante salientar ainda, que as mesmas variáveis de controle, de resultado e tratamento serão utilizadas na estimação do efeito por quantis. O estimador do efeito quantílico do tratamento pode ser usado com regressores endógenos, quando tiver instrumentos disponíveis, ou pode usar o status do tratamento exógeno, sendo reduzida ao estimador de regressão quantílica padrão.

Para a descrição do estimador, tem-se que os quantis condicionais dos resultados potenciais para os *compliers*<sup>20</sup> são dados por  $Q_\tau = (Y_i^0 | X_i, S_i) = X' \beta_\tau$  e  $Q_\tau = (Y_i^1 | X_i, S_i) = \alpha_\tau + X' \beta_\tau$ , onde  $\tau$  refere-se ao  $\tau$ -ésimo quantil pertencente ao intervalo (0,1). Assim sendo, dada a variação exógena do tratamento nos resultados potenciais, a função quantílica condicional pode ser escrita como

$$Q_\tau = (Y | X_i, S_i, S_1 > S_0) = \alpha_\tau S + X' \beta_\tau. \quad (5)$$

O  $\alpha_\tau$  pode representar uma relação causal dado que mostra a diferença entre o  $\tau$ -ésimo quantil dos resultados potenciais ( $Y_i^1$  e  $Y_i^0$ ) para os grupos de tratamento e controle, respectivamente. Em outras palavras, o coeficiente  $\alpha_\tau$  indica se houve mudança no rendimento do trabalho em decorrência da condição de saúde individual.

Como no efeito médio, o tratamento também pode ser correlacionado com o termo de erro o que produziria estimativas viesadas. Numa primeira estimativa assume-se que tanto o tratamento  $S$ , quanto  $X$  são exógenas, ou seja,  $\varepsilon \perp\!\!\!\perp (S, X)$ <sup>21</sup>. Dessa forma tem-se que  $Q_{y|x,s}^\tau = X' \beta_\tau + \alpha_\tau S$ , tal que é possível recuperar os parâmetros desconhecidos dos resultados potenciais das distribuição conjunta de  $y$ ,  $X$  e  $S$ , e a

<sup>16</sup> Hipótese de independência condicional. Para mais detalhes ver (Rubin, 1974; Heckman e Robb Jr, 1985).

<sup>17</sup> Do inglês *Average Treatment Effect on the Treated*.

<sup>18</sup> A subseção 4.1.1 apresenta o método usado para corrigir o viés de seleção, tanto no caso do tratamento médio como no efeito de tratamento quantílico.

<sup>19</sup> Do inglês *quantile treatment effects*.

<sup>20</sup> São os agrupamentos de acordo com o status de tratamento.

<sup>21</sup> A seção de robustez será responsável por respaldar ou não tais estimativas.

estimação pode ser feita pela regressão quantílica clássica como sugerida por Koenker e Bassett (1978), tal como

$$(\hat{\beta}^\tau, \hat{\delta}^\tau) = \operatorname{argmin}_{\beta, \delta} \sum W_i \times \rho_\tau(y_i - X_i\beta - H_i\delta) \quad (6)$$

onde  $\rho_\tau(u) = u \times \{\tau - 1(u < 0)\}$  e  $W_i$  são os pesos. Tal estimador produz erros padrão analíticos que são consistentes também na presença de heterocedasticidade.

As próximas duas seções apresentam métodos de estimações robustos para contribuir com os achados desta seção. Antes será apresentado o método de correção do viés de seleção amostral, usado tanto para o estimador do efeito médio quanto para o estimador quantílico.

#### 4.1.1 Método de Correção do Viés de Seleção

A renda do trabalho é observada somente para aqueles indivíduos que estão no mercado de trabalho. Tal fato conduz a auto-seleção da amostra e a não consideração de características que fazem com que o indivíduo decida por ofertar trabalho pode viesar as estimativas. Para obter a probabilidade de trabalhar ou não trabalhar existem diversos métodos, sendo o mais conhecido o *Heckit*, que usa um probit para calcular a inversa da razão de Mills e corrigir o viés de seleção (Heckman, 1979).

No entanto no presente trabalho será utilizado o método proposto por De Luca e Perotti (2011) e aplicado por Oliveira (2014), que considera que a probabilidade de trabalhar ou não segue um processo não paramétrico. Dessa forma, Gallant e Nychka (1987) demonstram que a densidade conjunta desconhecida de  $f$  pode ser aproximada por uma expansão polinomial de Hermite, sendo dada por

$$f^*(u_1, u_2; \gamma) = \frac{1}{\psi_R(\gamma)} \tau_R(u_1, u_1; \gamma)^2 \phi(u_1) \phi(u_2), \quad (7)$$

onde  $\gamma$  é um vetor dos parâmetros desconhecidos,  $\tau_R(u_1, u_1; \gamma)^2$  é um polinômio de ordem  $R$  em  $u_1$  e  $u_2$ ,  $\phi$  é uma função de densidade normal padrão e  $\psi_R(\gamma)$  é uma normalização para garantir que  $f^*$  seja uma função de densidade. Integrando a equação (7), a função de densidade conjunta pode ser aproximada por

$$\begin{aligned} F^*(u_1, u_2; \gamma) = & \Phi(u_1)\Phi(u_2) + \frac{1}{\psi_R(\gamma)} A_{12}^*(u_1, u_2; \gamma) \phi(u_1) \phi(u_2) \\ & - \frac{1}{\psi_R(\gamma)} A_1^*(u_1; \gamma) \Phi(u_2) \phi(u_1) - \frac{1}{\psi_R(\gamma)} A_2^*(u_2; \gamma) \Phi(u_1) \phi(u_2), \end{aligned} \quad (8)$$

em que  $A_{12}^*(u_1, u_2; \gamma)$ ,  $A_1^*(u_1; \gamma)$  e  $A_2^*(u_2; \gamma)$  são polinômios em  $u_1$  e em  $u_2$ . As distribuições marginais de  $u_1$  e  $u_2$  podem ser encontradas de forma semelhantes<sup>22</sup>. Depois disso, o vetor de parâmetros é estimado de maneira semiparamétrica a partir da maximização da função de pseudoverossimilhança.

O vetor de variáveis explicativas incluído nessa etapa é formado, em especial, por covariadas sobre a decisão de oferta de trabalho, incluindo a renda do não trabalho (aluguel, juros, pensão, aposentadoria, etc), características pessoais e familiares, hábitos de saúde, prática de exercícios e dificuldades de mobilidade/locomoção. Todas as variáveis utilizadas nesse estágio estão também descritas na tabela A.1, em anexo.

A variável dependente é a participação ou não no mercado de trabalho, e segundo Buchinsky (1998) deve ser estimada por um modelo semiparamétrico, por isso a escolha do modelo de resposta binária de estimadores de máxima verossimilhança semi não-paramétricos<sup>23</sup> (SNP) de Gallant & Nychka (1987). Dessa etapa é obtida a inversa da razão de Mills, que junto com seu quadrado é inserida como variável explicativa na estimação do efeito de tratamento médio e quantílico, para avaliar a presença ou não de viés na seleção amostral.

#### 4.2 Identificação com Heterocedasticidade

Dada à ausência de variáveis instrumentais externas que possam ser usadas diretamente nos modelos estimados na subseção anterior, a presente seção descreve o estimador de dois estágios proposto por Lewbel

<sup>22</sup> Para mais detalhes ver De Luca (2008) e Oliveira (2014).

<sup>23</sup> O comando semi-nonparametric (SNP) desenvolvido por De Luca (2008) é usado para a estimação desses coeficientes.

(2012), que explora a heterocedasticidade do primeiro estágio da regressão para gerar instrumentos internamente que possibilitem a identificação, quando não existem restrições de exclusão disponíveis<sup>24</sup>.

Seguindo Lewbel (2012), o sistema triangular de equações como apresentado abaixo mostra que quando as correlações dos erros são causadas devido a fatores não observados, a identificação se dá tendo regressores não correlacionados com o produto dos erros heterocedásticos.

Para o propósito desse artigo, supondo  $U$  ser uma variável omitida que possivelmente afeta a variável endógena de tratamento “Estado de saúde” (denotada por  $S$ ) tanto quanto afeta o resultado renda do trabalho  $Y$ , e denotando  $V_1$  e  $V_2$  como erros idiossincráticos. Portanto, é possível identificar o efeito causal das condições de saúde em  $Y$ , denotado por  $\beta$ , através do Método dos Momentos Generalizados (GMM) ou de Mínimos Quadrados Modificados em Dois Estágios (2SLS).

$$Y = X'\beta_1 + \beta S + \varepsilon_1 \quad (9)$$

$$S = X'\beta_2 + \varepsilon_2 \quad (10)$$

$$E[X\varepsilon_1] = 0, \quad E[X\varepsilon_2] = 0, \quad Cov[Z, \varepsilon_1\varepsilon_2] = 0, \quad (11)$$

Onde  $\varepsilon_1 = \alpha_1 U + V_1$  e  $\varepsilon_2 = \alpha_2 U + V_2$  e  $Z \subseteq X$ <sup>25</sup>. A equação (11) apresenta as condições necessárias para a identificação e estimação, juntamente com alguma heterocedasticidade em  $\varepsilon_i$ <sup>26</sup>. Assim, a classe de modelos que satisfazem os pressupostos subjacentes ao método de Lewbel são aqueles que as correlações dos erros nas equações em *cross-section* são por conta da presença de um fator comum não observado.

A equação auxiliar ou regressão do primeiro estágio pode ser usada para fornecer os elementos necessários para o método de Lewbel. Em sua versão mais simples, instrumentos podem ser construídos a partir dos resíduos das regressões auxiliares multiplicado por cada uma das variáveis exógenas centrada na média, ou seja:

$$Z_j = (X_j - \bar{X}) \cdot \epsilon \quad (12)$$

onde  $\epsilon$  é o vetor de resíduos da regressão do primeiro estágio de cada regressor endógeno contra todos os regressores exógenos, incluindo um vetor de constantes. Esses resíduos tem covariância zero com cada um dos regressores usados para construí-lo, o que implica que a média dos instrumentos gerados são zero. Por outro lado, o produto desses resíduos com os elementos dos regressores centrados na média não será zero, se houver evidência considerável de heterocedasticidade de escala com relação às variáveis explicativas.<sup>27</sup> Assim, quanto maior o grau da heterocedasticidade no processo de erro, maior será a correlação dos instrumentos gerados com as variáveis endógenas incluídas (que são os regressandos das regressões auxiliares).

É claro que o método fornece estimativas menos confiáveis do que se tivesse uma variável instrumental exógena para ser usada nas estimativas. No entanto, alguns trabalhos empíricos como Erman e Hou (2013), Amorim e Sampaio (2015) e Tigre *et al.* (2016) tem mostrado que a abordagem de Lewbel produz resultados muito parecidos quando comparado com a estimativa obtida com o uso de instrumentos tradicionais.

Em suma, na falta de uma variável instrumental externa válida, como é o caso de muitas aplicações empíricas, o método de Lewbel é uma alternativa interessante para estimar o efeito causal de interesse, nesse caso do estado de saúde dos indivíduos sobre o salário por hora dos indivíduos.

#### 4.3 Bounding para o Efeito de Tratamento

Como último procedimento empírico adotado no presente trabalho, esta seção apresenta o método recentemente proposto por Oster (2013; 2015) para estimar bounds para o efeito de tratamento, usado para corroborar os resultados encontrados com as estratégias das subseções anteriores.

A ideia é que os movimentos no coeficiente de interesse, nesse caso a condição de saúde individual, carrega informação sobre o viés remanescente devido a variáveis não observadas, dada a inclusão ou não de controles devido à preocupação com variáveis omitidas. Em outras palavras, o método segue a noção de

<sup>24</sup> O método pode ser usado também para melhorar a eficiência do estimador IV padrão.

<sup>25</sup> Note que  $Z$  é um subconjunto de  $X$ , e com isso, nenhuma informação de fora do modelo especificado é necessária.

<sup>26</sup> Para mais detalhes ver Lewbel (2012) e Baum e Schaffer (2012).

<sup>27</sup> A heterocedasticidade pode ser analisada por testes tal como o de Breusch-Pagan.

Altonji et al (2005) que as variáveis não observáveis não devem ser mais importantes do que as observáveis na explicação do tratamento.

Para descrever o método, considere que existe um conjunto de variáveis não observáveis,  $U$ , tal como:

$$Y = \alpha + \beta S + X'\gamma + U + \varepsilon \quad (13)$$

O pressuposto de seleção proporcional afirma que  $\delta \frac{\sigma_{XS}}{\sigma_X} = \frac{\sigma_{US}}{\sigma_U}$ , onde  $\sigma_{XS} = \text{cov}(X, S)$ ,  $\sigma_{US} = \text{cov}(U, S)$ ,  $\sigma_X = \text{Var}(X)$  e  $\sigma_U = \text{Var}(U)$ , e o coeficiente de proporcionalidade,  $\delta$ , é informativo sobre como as não observáveis se relacionam com o tratamento, dado que é conhecido como as observáveis se relacionam com tal tratamento. Dessa forma, se  $\delta = 1$ , por exemplo, a seleção em não observáveis é igual à seleção sobre as observáveis.

Dito isso, considere as equações a seguir, onde o R-quadrado do modelo completo representado pela Equação (14) é o  $R_{max}$ , e para as equações (15) e (16) é  $R_1$  e  $R_2$ , respectivamente. A Equação (16) inclui  $M$ , que é um conjunto restrito de controles observados e não tem um componente não observado associado, além de ser ortogonal com  $X$  e  $U$ .

$$y = \alpha + \beta S + X'\gamma + U + \varepsilon_{max} \quad (14)$$

$$y = \alpha + \beta_1 S + X'\gamma + \varepsilon_1 \quad (15)$$

$$y = \alpha + \beta_2 S + M'\alpha + \varepsilon_2 \quad (16)$$

Sob essas restrições, e quando  $\delta$  é próximo de 1, o valor de  $B(\delta) = \delta \frac{(\beta_2 - \beta_1)(R_{max} - R_1)}{R_1 - R_2}$  é igual a:

- a) O viés não observado se  $\delta = 1$ ;
- b) Próximo ao limite superior do viés se  $\delta < 1$ ;
- c) Próximo ao limite inferior do viés se  $\delta > 1$ ;

Observe que o vetor de controles observados  $X$ , é uma seleção aleatória do conjunto total  $(X, U)$ , quando  $\delta = 1$ . Oster argumenta que as covariáveis mais importantes para explicar o tratamento geralmente estão contidas no vetor  $X$ , o que implica que  $\delta$  deve ser maior que um. Em outras palavras, a seleção em não observáveis não deve exceder a seleção sobre as observáveis.

Portanto, dada essa configuração, é possível obter estimativas não viesadas para o efeito de tratamento, calculando o viés via  $B(\delta)$ , que para seu cálculo é necessário conhecer o valor de  $R_{max}$ , que não é observado. No entanto, deve haver alguma aleatoriedade nos movimentos da variável de resultado fazendo com que o  $R_{max}$  seja menor que 1. Nas estimações desse presente trabalho foram considerados valores para o  $R_{max}$  variando de 0,7<sup>28</sup> até 1, sendo este último um valor muito prudente, usado para o cálculo de  $\delta$ .<sup>29</sup>

De acordo com a abordagem efetuada em Altonji et al (2005) e Oster (2013), é possível obter um valor de  $\delta$  que seria suficiente para explicar o efeito de tratamento por completo, isto é, que faria com que  $\hat{\beta} = 0$ . Esse valor daria uma ideia do grau de seleção nas não observáveis relativo às observáveis que seria necessário para o tratamento ser completamente explicado por variáveis não observadas, não incluídas no modelo.

## 5. Resultados

Esta seção apresenta o efeito das condições de saúde sobre os rendimentos do trabalho com base no *propensity score* para o efeito de tratamento médio e o efeito de tratamento quantílico, usando características pessoais, de localização, de ramo de atividade e *status* de trabalho como controles. Na subseção 5.1 serão apresentados os resultados baseados nos métodos de Lewbel (2012) e Oster (2015) como robustez para as estimativas aqui apresentadas.

Antes disso, a Tabela 4 apresenta as variáveis que explicam as diferenças de probabilidade de ter uma má condição de saúde autoreportada ou ser diagnosticado com alguma doença por um profissional,

<sup>28</sup> Usado em Oster (2013)

<sup>29</sup> Também usado em Oliveira et al (2015). Para mais detalhes ver Oster (2013).

através de um *propensity score* logístico. Foram incluídas todas as características observadas que podem afetar tanto o tratamento (condição de saúde) quanto o resultado (renda do trabalho). É importante lembrar que as variáveis binárias das condições de saúde são construídas de modo a assumir o valor 1 para uma má condição autoreportada (Autoavaliação de Saúde) e para ter sido diagnosticado com alguma doença crônica (Doenças). Dessa forma coeficientes positivos indicam que a variável em questão impacta para aumentar a péssima condição de saúde enquanto que coeficientes negativos impactam em reduzir a má condição de saúde.

Apesar de uma diferença no valor dos coeficientes entre as duas variáveis de saúde, ser do sexo feminino e casado afetam a probabilidade de ter uma má condição de saúde. Como visto na seção 3.1, um maior percentual de mulheres é diagnosticado com praticamente todas as doenças listadas em relação aos homens assim como uma maior proporção relatam uma má condição de saúde autoavaliada. Tal resultado segue em linha com Gomes et al (2012), e pode ser explicado pelo fato das mulheres fazerem mais consultas médicas (inclusive exames de rotina e prevenção) e com isso conhecerem melhor sua real condição.

Outras variáveis que contribuem para uma maior probabilidade de saúde insatisfatória é o aumento da idade, que naturalmente debilita a saúde individual, e do tamanho da família, que faz com que se tenha menos recurso *per capita* para cuidados com a saúde. Ser fumante e residir em área urbana também aumenta a probabilidade de não ser saudável, esta última possivelmente pelo estilo de vida mais estressante levado nas cidades, apesar do maior acesso aos serviços de saúde.

**Tabela 4: Matching Logit: Regressão do Tratamento em todas as outras variáveis**

Variáveis	Tratamento			
	Auto Avaliação de Saúde	Desvio Padrão	Doenças <sup>a</sup>	Desvio Padrão
Sexo	0,3477***	(0,0253)	0,5311***	(0,0216)
Raça	-0,2143***	(0,0231)	-0,0627***	(0,0201)
Estado Civil	0,1162***	(0,0256)	0,0901***	(0,0224)
Idade	0,0380***	(0,0010)	0,0599***	(0,0009)
Ano de Estudo	-0,0815***	(0,0029)	-0,0259***	(0,0025)
Fumante	0,0523**	(0,0251)	0,0649***	(0,0226)
Atividade Física	-0,3659***	(0,0257)	-0,0880***	(0,0210)
Renda (Não trabalho) <sup>1</sup>	-0,0001	(0,0001)	-0,0001	(0,0001)
Tamanho da família	0,0243***	(0,0074)	-0,0115*	(0,0067)
Militar <sup>2</sup>	-0,0095	(0,0518)	0,1137***	(0,0428)
Com Carteira	-0,1389***	(0,0343)	0,0267	(0,0299)
Doméstico	0,2720*	(0,1407)	0,0127	(0,1117)
Conta – Própria	0,1250***	(0,0331)	0,1296***	(0,0302)
Empregador	-0,2457***	(0,0568)	0,1309***	(0,0463)
Urbana	0,0461	(0,0353)	0,1257***	(0,0330)
Constante	5,6249***	(0,3582)	4,7451***	(0,4025)
Ocupação (FE)	Sim		Sim	
UF (FE)	Sim		Sim	
Observações	60.812		60.812	
Log likelihood	-38.029,045		-45.886,874	

Fonte: Elaboração própria a partir dos dados das PNAD de 2008.

Notas: Desvios-padrão entre parênteses. \*\*\*p-valor < 0,01. \*\* p-valor < 0,05. \* p-valor < 0,10

<sup>a</sup> Ter pelo menos uma doença constatada por um médico.

<sup>1</sup> Soma de todos os rendimentos do não trabalho (pensão, aposentadorias, juros, aluguel, etc); <sup>2</sup> Categoria base: sem carteira; <sup>3</sup> Categoria base: Região Sudeste

Por outro lado, indivíduos brancos e que praticam atividade física reduzem a probabilidade de reportar uma condição não saudável, sendo que a raça pode estar ligada ao diferencial de rendimentos em relação aos não-brancos, fato bastante consolidado na literatura sobre discriminação salarial (Ver por exemplo Cacciamali e Hirata, 2005 e Coelho et al, 2010). O nível de escolaridade também contribui para uma menor probabilidade de má saúde, resultado consolidado na literatura que mostra uma relação positiva entre educação e *status* de saúde, tal como em Sousa (2010).

A relação entre educação e saúde pode se dar diretamente, ou a partir dos efeitos da renda. De maneira direta, pessoas mais escolarizadas tendem a adotar hábitos de vida mais saudáveis e a procurar mais

os serviços médicos, especialmente os cuidados preventivos, uma vez que, ao possuírem mais informações sobre as doenças e os possíveis tratamentos ou diagnósticos, tendem a valorizar de forma diferenciada o cuidado com a saúde. Com relação à renda, esta tende a aumentar com o nível de escolaridade, de modo que o indivíduo dispõe de melhores condições de vida e de acesso aos serviços de saúde.

Até esse ponto os resultados coincidem para as duas variáveis de saúde. No entanto, para o ramo de atividade algumas características têm impactos contrários, destacando o fato de ser empregador que reduz a probabilidade da autopercepção de má saúde e aumenta a probabilidade de ter alguma doença diagnosticada. Isso pode ser devido à carga de trabalho, que acarreta em alguma doença crônica diagnosticada. Por fim, ainda foram incluídos *dummies* de efeito fixo para cada um dos Estados e para o tipo de ocupação no mercado de trabalho.

A Tabela 5 abaixo apresenta o efeito médio com a correção do viés de seleção amostral para diferentes condições de saúde usadas como tratamento. Além da condição autoavaliada e do diagnóstico de pelo menos uma doença, também são usados o diagnóstico de duas ou mais doenças e de três ou mais doenças, como já explicitado antes, sendo o grupo de controle o indivíduo saudável em ambos os casos.<sup>30</sup>

O efeito é negativo em todas as especificações, evidenciando uma redução de rendimentos do trabalho devido a uma condição de saúde insatisfatória. O impacto da autoavaliação de saúde é maior que o diagnóstico de uma doença, o que mesmo sendo um critério subjetivo, mensura de forma mais adequada o estado de saúde e evidencia que indivíduos que afirmam ter uma má condição de saúde autoreportada tendem a ter um menor desempenho e produtividade, comparado ao diagnóstico de uma doença.

À medida que o número de doenças usado como tratamento aumenta, é observada uma elevação na magnitude do coeficiente, indicando uma deterioração da saúde e consequentemente da capacidade produtiva individual, refletindo assim no aumento da chance de rendimento do trabalho.

Este resultado caminha no mesmo sentido das evidências empíricas presentes na literatura que mostram um efeito negativo da condição de saúde no rendimento, nas horas trabalhadas ou até mesmo na exclusão do mercado de trabalho. No entanto, foram utilizados aqui vários tratamentos para evidenciar a importância da saúde autoavaliada como reconhecimento do indivíduo acerca da sua capacidade e ainda aumento do efeito quando cresce o número de doenças diagnosticadas no indivíduo.

**Tabela 5: Estimação do Efeito de Tratamento Médio**

Método	Tratamento			
	Autoavaliação	Doenças (1 doença)	Doenças (2 ou mais)	Doenças (3 ou mais)
ATT	-0,117*** (0,006)	-0,042*** (0,013)	-0,356*** (0,019)	-0,799*** (0,050)
Controles	Sim	Sim	Sim	Sim
Ocupação (FE)	Sim	Sim	Sim	Sim
UF (FE)	Sim	Sim	Sim	Sim
Observações	60.832	60.832	46.397	40.735

Fonte: Elaboração própria a partir dos dados das PNAD de 2008.

Notas: Desvios-padrão robustos a heterocedasticidade entre parênteses, com bootstrap 1000 replicações.

\*\*\*p-valor < 0,01. \*\* p-valor < 0,05. \* p-valor < 0,10

Foi também utilizada cada uma das doze doenças listadas na seção de dados como tratamento individual, controlado por todas as demais características supracitadas, inclusive pelas demais doenças. Com exceção do câncer e cirrose, que apesar de apresentar um coeficiente positivo não foi estatisticamente significativo, todas as demais doenças impactaram negativamente no salário por hora dos trabalhadores, chegando a -0,209 o efeito da doença renal<sup>31</sup>.

Para verificar o impacto ao longo de toda a distribuição de rendimentos, a Figura 1 sintetiza os resultados do efeito de tratamento quantílico corrigido pelo viés de seleção para cada um dos quatro

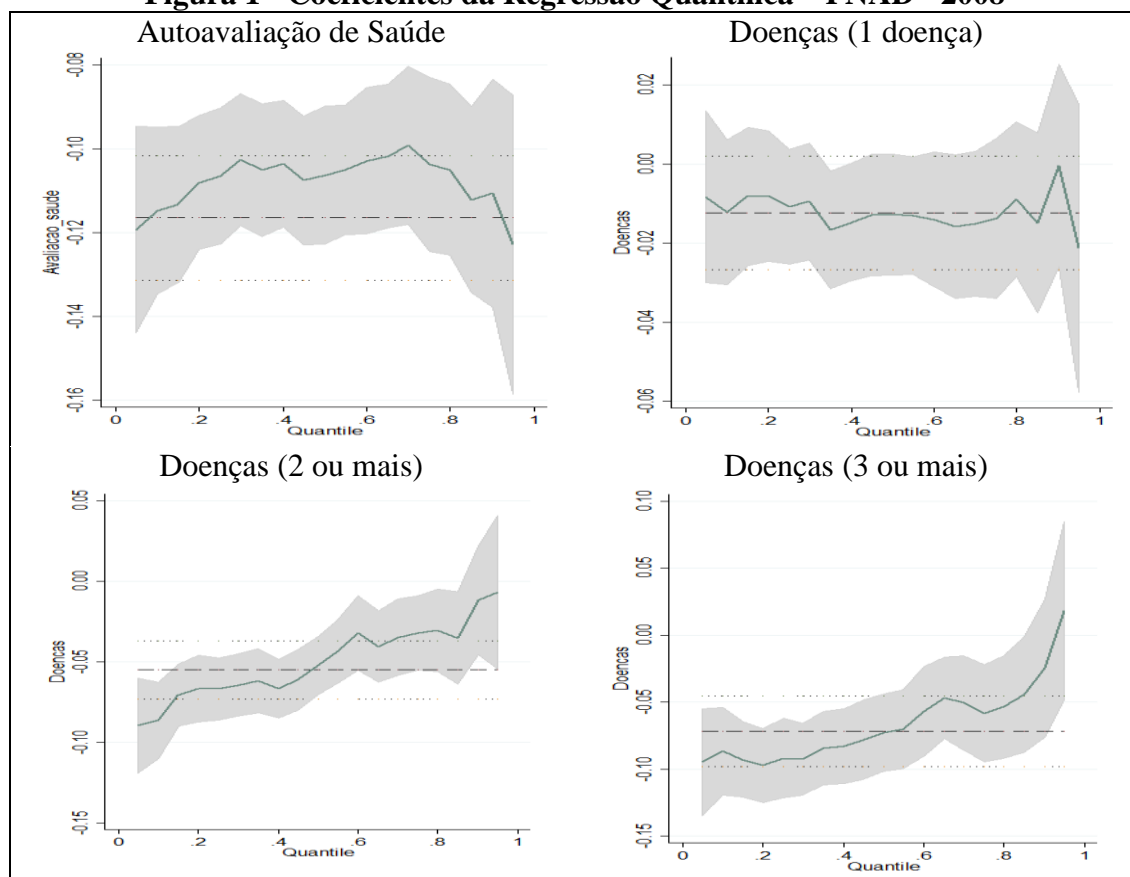
<sup>30</sup> Por exemplo, na construção da *dummy* que indica ter duas ou mais doenças = 1 e nenhuma doença = 0 os indivíduos que apresentam apenas uma doença são excluídos. O mesmo é feito para a *dummy* que indica ser diagnosticado com três ou mais doenças.

<sup>31</sup> Os resultados não são apresentados por falta de espaço, mas podem ser solicitados aos autores.

tratamentos utilizados. Diferente da estimação pela média condicional da variável dependente, este método permite visualizar como os efeitos do tratamento se alteram em diferentes quantis.

São plotadas na Figura as estimativas do coeficiente da *dummy* de condição de saúde para cada quantil entre 0,05 e 0,95, bem como o intervalo de confiança de 95%, representado pela área cinza em volta da curva. Assim como no efeito médio, a redução analisada pela autoavaliação de saúde é maior que o diagnostico de uma doença ao longo de toda a distribuição, corroborando com a ideia de um indicador global da real condição de saúde para este primeiro. O impacto maior da autoavaliação de saúde se da nos dois extremos da distribuição, o que pode indicar uma má percepção pela condição de pobreza nos quantis inferiores, e uma carga excessiva de trabalho nos quantis superiores. Com relação ao diagnostico de uma doença, os coeficientes mantém certa estabilidade ao longo da distribuição, sendo o efeito negativo, mas próximo de zero.

**Figura 1 - Coeficientes da Regressão Quantílica – PNAD - 2008**



Fonte: Elaboração própria com base nas estimações.

Os coeficientes ao longo de toda a distribuição para o tratamento duas ou mais doenças e três ou mais doenças se mostraram menores que aquele do efeito médio, o que sugere evidências de uma superestimação do impacto na média. Além disso, os trabalhadores localizados nos menores quantis tiveram uma redução de salário muito maior que os quantis superiores, corroborando a ideia do círculo vicioso como visto em Gomes et al (2012). Isso significa que os trabalhadores que tem menores rendimentos são justamente aqueles que têm a maior perda quando se encontram doentes, o que pode torna-los mais pobres e agravar mais ainda a condição de saúde. Além disso, os trabalhadores com menores rendimentos são aqueles que tendem a ocupar trabalhos que demandam maior esforço físico, ficando muitas vezes impossibilitado de exercer suas tarefas com alguma enfermidade.

O efeito negativo vai se tornando menor ao longo da distribuição chegando a ser praticamente zero para os indivíduos que recebem os mais altos rendimentos. Em suma, o efeito de uma má condição de saúde é menor para aqueles indivíduos que tem maiores rendimentos, o que implica uma maior capacidade de adaptação desses trabalhadores ao mercado de trabalho, levando em conta a gravidade da enfermidade e o fato de que em geral ocupam trabalhos que exigem menos esforço físico, sendo possível conviver com tal condição.

### 5.1 Análise de Robustez

As evidências mostram o impacto negativo de uma má condição de saúde na renda do trabalho individual. Esta seção começa pela apresentação dos resultados do efeito de tratamento médio estimado por *Generalized Method of Moments* (GMM) a partir do método de Lewbel (2012), que explora a identificação a partir da heterocedasticidade dos resíduos para gerar instrumentos internos, quando o produto de tais erros não é correlacionado com os regressores.<sup>32</sup>

As estimativas corroboram os resultados anteriores de um efeito negativo da condição de saúde sobre o rendimento dos trabalhadores, independente do tratamento utilizado. No entanto, tal como no efeito quantílico, o método de Lewbel evidencia uma possível superestimação do efeito médio quando não considerados os instrumentos para a condição de saúde. Tal fato pode ser visto pela redução de todos os coeficientes, como por exemplo da saúde autoavaliada como tratamento, que passa de -0,117 (efeito de tratamento médio) para -0,0619 (efeito médio pelo método de Lewbel). Para duas/três ou mais doenças como tratamento, a redução do coeficiente é ainda maior.

A última linha da Tabela 6 apresenta o teste de subidentificação do modelo, o qual se rejeita a hipótese nula indicando que o modelo é identificado. Já o teste de Cragg-Donald rejeita a hipótese nula de fracos instrumentos, sugerindo que os instrumentos internos gerados são relevantes para o tratamento<sup>33</sup>. O artigo de Emran e Hou (2013) encontra uma semelhança qualitativa entre as estimativas obtidas por instrumentos externos e pelos gerados pelo presente método, de modo que apesar de não dispor de variáveis de exclusão externas é possível ter certo grau de confiança nos instrumentos gerados para o presente trabalho.

**Tabela 6: ATT Estimado usando o Método de Lewbel (2012)**

Método	Tratamento			
	Auto Avaliação	Doenças (1 doença)	Doenças (2 ou mais)	Doenças (3 ou mais)
ATT <sub>GMM</sub>	-0,0619*** (0,0175)	-0,0318** (0,0190)	-0,0720*** (0,0161)	-0,1070** (0,0464)
Controles	Sim	Sim	Sim	Sim
Ocupação (FE)	Sim	Sim	Sim	Sim
UF (FE)	Sim	Sim	Sim	Sim
Observações	60.548	60.548	46.174	40.531
Kleibergen-Paap LM	4017.766***	47.591***	95.640***	1497.596***

Fonte: Elaboração própria a partir dos dados das PNAD de 2008.

Notas: Desvios-padrão robustos a heterocedasticidade entre parênteses, com bootstrap 1000 replicações.

\*\*\*p-valor < 0,01. \*\* p-valor < 0,05. \* p-valor < 0,10

Por fim, são estimados os *bounds* para o efeito de tratamento da condição de saúde seguindo Oster (2015), para obter os valores da seleção nas não observáveis que seria necessário para explicar completamente as estimativas, em comparação com a seleção nas observáveis. Para cada variável de tratamento são considerados quatro valores de  $R_{max}$  para se obter o valor do coeficiente de proporcionalidade ( $\delta$ ) que seria necessário para fazer a estimativa do efeito de tratamento ir para zero ( $\beta=0$ ) e também para se obter um limite inferior para o efeito (o conjunto identificado).

A Tabela 7 apresenta tais resultados para cada condição de saúde usada como tratamento. Para a autoavaliação de saúde, considerando o  $R_{max}$  de 0,7, por exemplo, foi encontrado um coeficiente de proporcionalidade de 2,8034, o que implica que o efeito das variáveis não observadas no ATT estimado ao longo da condição de saúde teria que ser 2,8034 vezes mais forte do que o efeito das variáveis observáveis para explicar todo o efeito negativo da condição de saúde no rendimento do trabalho. Em outras palavras, as não observáveis teriam que ser 2,8034 vezes mais importantes na explicação do efeito médio, de modo a fazer com que o coeficiente baseado nas observáveis fosse zero. Aumentando o  $R_{max}$  para um valor muito

<sup>32</sup> Foi realizado o teste de White (1980) e rejeitada a hipótese nula de homocedasticidade (verificada a presença de heterocedasticidade), condição para a validade do método.

<sup>33</sup> Os resultados do teste podem ser solicitados junto aos autores.



improvável de 1, tal coeficiente ainda é de 1,4694. Para o limite inferior, o valor é de -0,20 (para  $R_{max} = 0,7$ ) e passa para -0,09 (para  $R_{max} = 1,0$ ).

**Tabela 7: Bounds para o Efeito Do Tratamento – Oster (2015)**

<b>Auto Avaliação</b>				
Parâmetro	$R_{max}$			
	0.7	0.8	0.9	1.0
$\delta$ para $\beta=0$	2,8034	2,1521	1,7464	1,4694
Identificado ( $\delta=1$ )	[-0,28;-0,20]	[-0,28;-0,16]	[-0,28;-0,13]	[-0,28;-0,09]
<b>Doenças (1 Doença)</b>				
Parâmetro	$R_{max}$			
	0.7	0.8	0.9	1.0
$\delta$ para $\beta=0$	2,2366	1,7202	1,3975	1,1768
Identificado ( $\delta=1$ )	[-0,28;-0,17]	[-0,28;-0,13]	[-0,28;-0,09]	[-0,28;-0,04]
<b>Doenças (2 ou mais)</b>				
Parâmetro	$R_{max}$			
	0.7	0.8	0.9	1.0
$\delta$ para $\beta=0$	2,3341	1,7972	1,4611	1,2309
Identificado ( $\delta=1$ )	[-0,27;-0,17]	[-0,27;-0,13]	[-0,27;-0,09]	[-0,27;-0,05]
<b>Doenças (3 ou mais)</b>				
Parâmetro	$R_{max}$			
	0.7	0.8	0.9	1.0
$\delta$ para $\beta=0$	2,4948	1,9111	1,5487	1,3019
Identificado ( $\delta=1$ )	[-0,27;-0,18]	[-0,27;-0,14]	[-0,27;-0,10]	[-0,27;-0,06]

Fonte: Elaboração própria com base nas estimativas

Para as demais variáveis de tratamento a interpretação é semelhante, o que corrobora os achados das tabelas anteriores, qual seja, de que o viés causado pelas não observáveis não inviabiliza os coeficientes estimados. Oster (2015) analisando o impacto do comportamento materno sobre as crianças encontrou um coeficiente de proporcionalidade,  $\delta$ , de 1,37 e o conjunto identificado de [-0,124;-0,033] que exclui o zero. Tal efeito mostrou-se significativo, inclusive quando comparado a outros métodos. Diante disso, tem-se uma confiabilidade nos resultados apresentados na presente pesquisa e corroborados nesta seção.

## 6. Considerações Finais

As condições de saúde, juntamente com a educação, são fatores primordiais para a acumulação de capital humano e consequentemente aumento da produtividade e aferição de renda. Ciente disso foi estimado o impacto da má condição de saúde nos rendimentos do trabalho com dados da PNAD de 2008 que contem um suplemento de saúde, a partir de quatro métodos empíricos: efeito de tratamento médio e efeito de tratamento quantílico, ambos com correção semiparamétrica para o viés de seleção amostral; estimação a partir de instrumentos internos gerados pela heterocedasticidade dos erros do primeiro estágio seguindo Lewbel (2012), e a estimação dos *bounds* para o efeito de tratamento de Oster (2015), juntamente com o coeficiente de proporcionalidade que verifica o viés causado pelas variáveis não observáveis.

Foram utilizadas quatro variáveis indicadoras do *status* de saúde como tratamento como forma de corroborar as estimativas. A condição de saúde autoavaliada, o diagnóstico de pelo menos uma doença por um profissional de saúde, o diagnóstico de duas ou mais doenças e o diagnóstico de três ou mais doenças. Os resultados mostraram um impacto negativo de uma má condição de saúde sobre os salários para todas as estimações, com uma possível superestimação do efeito médio quando comparado ao método de instrumentos internos de Lewbel (2012).

Quando a análise é estendida ao longo da distribuição de renda, é verificado que os indivíduos localizados nos quantis inferiores sofrem uma maior redução salarial devido a uma condição de saúde insatisfatória, o que pode agravar mais ainda a precariedade da saúde desses indivíduos, criando um “círculo vicioso” em tal relação, dado que os trabalhos que remuneram menos são em geral os que demandam maior

esforço físico e consequentemente os mais afetados e de difícil adaptação por problemas de saúde, dada a impossibilidade de empreender o mesmo nível de esforço caso estivesse saudável.

Por fim, o coeficiente de proporcionalidade, segundo Oster (2015), corrobora com os resultados anteriores e mostra que o viés causado pela omissão de variáveis ou por variáveis não observáveis não é capaz de tornar as estimativas insignificantes, servindo de robustez às estimações anteriores. Em suma, a má condição de saúde causa perdas salariais a todos os trabalhadores, sobretudo àqueles com menores rendimentos. Tais evidências reforçam a necessidade de mais oportunidades na área de saúde para esses trabalhadores, inclusive com prevenção e consultas/exames médicos de rotinas a fim de antecipar e tratar possíveis problemas e evitar uma baixa produtividade no mercado de trabalho.

## Referências

- ALTONJI, J., ELDER, T. e TABER, C. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. **Journal of Political Economy**, 113 (1): 151-184, 2005.
- ALVES, L. F. **Impactos do estado de saúde sobre os rendimentos individuais no Brasil**. Dissertação de Mestrado, 2002. Disponível em: <[http://www.cedeplar.ufmg.br/economia/dissertacoes/2002/Luiz\\_Fernando\\_Alves.pdf](http://www.cedeplar.ufmg.br/economia/dissertacoes/2002/Luiz_Fernando_Alves.pdf)>.
- ALVES, L. F.; ANDRADE, M. V. Impactos da saúde nos rendimentos individuais no Brasil. **Economia Aplicada**, v. 7, n. 2, p. 259–388, 2003.
- AMORIM, G., e SAMPAIO, B. **Communication Network and Civil Unrest: An Investigation of the Occupy Movement in the United States**. 37th Meeting of the Brazilian Econometric Society, Brazil, 2015.
- ANGRIST, J. D.; PISCHKE, J. **Mostly harmless econometrics: An empiricist's companion**. Princeton university press, 2009.
- BARTEL, A.; TAUBMAN, P. Health and Labor Market Success: The Role of Various Diseases. **The Review of Economics and Statistics**, v. 61, n. 1, p. 1–8, 1979.
- BAUM, C. F., e SCHAFFER, M.E. **ivreg2h: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments**. 2012. Disponível em: <<http://ideas.repec.org/c/boc/bocode/s457555.html>>.
- BOCKERMAN, P.; HYYTINEN, A.; KAPRIO, J. Smoking and long-term labour market outcomes. **Tobacco control**, p. 1–7, fev. 2014. ISSN 1468-3318. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24570102>>.
- BUCHINSKY, M. The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. **Journal of Applied Econometrics**, v. 13, n. 1, p. 1–30, 1998.
- CACCIAMALI, M. C.; HIRATA, G. I. A Influência da Raça e do Gênero nas Oportunidades de Obtenção de Renda – Uma Análise da Discriminação em Mercados de Trabalho Distintos: Bahia e São Paulo. **Estudos Econômicos**, São Paulo, v.35, n.4, p.767- 795, out-dez, 2005.
- CAI, L. The relationship between health and labour force participation: Evidence from a panel data simultaneous equation model. **Labour Economics**, 17(1): 77-90, 2010.
- COELHO, D.; VESZTEG, R.; SOARES, F. V. **Regressão quantílica com correção para a seletividade amostral: estimativa dos retornos educacionais e diferenciais raciais na distribuição de salários das mulheres no Brasil**. Brasília, DF: IPEA, 2010. (Texto para Discussão, n. 1483).
- DE LUCA, G. SNP and SML estimation of univariate and bivariate binary choice models. **The Stata Journal**, v. 8, p. 190-220, 2008.
- DE LUCA, G.; PEROTTI, V. Estimation of ordered response models with sample selection. **The Stata Journal**, v. 11, p. 213-39, 2011.
- DISNEY, R.; EMMERSON, C.; WAKEFIELD, M. Ill health and retirement in Britain: A panel data-based analysis. **Journal of Health Economics**, 25(4): 621-649, 2006.
- economics of health and medical care. London: Macmillan Press, p. 205-233, 1974.
- EMRAN, M. S., e HOU, Z. Access to Markets and Rural Poverty: Evidence from Household Consumption in China. **The Review of Economics and Statistics**, 95 (2): 682-697, 2013.
- FIGUEIREDO, L.; NORONHA, K. V. M. S.; ANDRADE, M. V. **Os impactos da saúde sobre o crescimento econômico na década de 90: uma análise para os estados brasileiros**. Belo Horizonte: CEDEPLAR/UFMG, 2003. 26 p. (Texto para Discussão nº 219).

GALLANT, A. R.; NYCHKA, D. W. Semi-Nonparametric Maximum Likelihood Estimation. **Econometrica**, v. 55, n. 2, p. 363–390, 1987.

GARCÍA-GÓMEZ, P.; JONES, A. M.; RICE, N. Health effects on labour market exits and entries. **Labour Economics**, 17(1): 62-76, 2010.

GODOY, M. R.; BALBINOTTO NETO, G.; RIBEIRO, E. P. **Estimando as Perdas de Rendimentos Devido a Doença Renal Crônica**. In: Encontro Nacional de Economia - ANPEC, 2006, Salvador. Anais do XXXIV Encontro Nacional de Economia -ANPEC, 2006.

GOMES, S. M. F. P. O.; BRITO, D. J. M.; ROCHA, R. M. **Impactos da Saúde sobre os Rendimentos Individuais no Brasil**. In: 40 ENCONTRO NACIONAL DE ECONOMIA, 2012, Porto de Galinhas, Pernambuco. Anais 40 ENCONTRO NACIONAL DE ECONOMIA, 2012.

GROSSMAN, M. On the Concept of Health Capital and the Demand for Health. **Journal of Political Economy**, v. 80, n. 2, p. 223–255, 1972.

GROSSMAN, M., BENHAM, L. **Health, hours and wages**. In.: PERLMAN, M. The

HECKMAN, J. J. Sample Selection Bias as a Specification Error. **Econometrica**, v. 47, n. 1, p. 153–161, 1979.

HECKMAN, J.J. e R. ROBB JR. Alternative methods for evaluating the impact of interventions: An overview. **Journal of Econometrics**, 30 (1-2): 239-267, 1985.

IMBENS, G. e J.M. WOOLDRIDGE. Recent Developments in the Econometrics of Program Evaluation. **Journal of Economic Literature**, 47 (1):5-86, 2009.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios**. Rio de Janeiro: IBGE, 2008. Disponível em: <<http://www.ibge.gov.br>>.

KASSOUF, A. L. Rendimentos perdidos por trabalhadores em condições inadequadas de saúde. **Economia Aplicada**. São Paulo, v. 3, n. 2, p. 239-262, 1999.

KASSOUF, A. L. Saúde e mercado de trabalho. **Pesquisa e Planejamento Econômico**, v. 27, n. 3, p. 587–610, 1997.

KOENKER, R. **Quantile Regression**. Cambridge: Cambridge University Press, 2005.

KOENKER, R.; BASSETT, G. Regression Quantiles. **Econometrica**, v. 46, n. 1, p. 33-50, 1978.

LEVINE, P. B.; GUSTAFSON, T. A.; VELENCHIK, A. D. More bad news for smokers? The effects of cigarette smoking on wages. **Industrial and Labor Relations Review**, v. 50, n. 3, p. 493–509, 1997.

LEWBEL, A. Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. **Journal of Business and Economic Statistics**, 30, 67-80, 2012.

LUFT, H.S. The impact of poor health on earnings. **The Review of Economics and Statistics**. Cambridge, v.57. 1975.

MINCER, J. **Schooling, experience and earnings**. New York: Columbia University Press, 1974.

MORRIS, S. The impact of obesity on employment. **Labour Economics**, 14(3): 413-433, 2007.

NORONHA, K; ANDRADE, M. **A importância da saúde como um dos determinantes da distribuição de rendimentos e pobreza no Brasil**. XXXII Encontro Nacional de Economia, Anpec, 2004.

OLIVEIRA, V. R.. Viés de seleção, migração e saúde. **Brazilian Journal of Theoretical and Applied Economics**, v. 20, p. 376-399, 2014.

OSTER, E. Unobservable selection and coefficient stability: theory and validation. NBER **working paper** No. w19054, 2013.

OSTER, E. Unobservable Selection and Coefficient Stability: Theory and Evidence. **Working paper**, Brown University, 2015.

ROSENBAUM, P. e RUBIN, D. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. **The American Statistician**, 39 (1): 33-38, 1985.

ROSENBAUM, P. e RUBIN, D. The central role of the propensity score in observational studies for causal effects. **Biometrika**, 70: 41-55, 1983.

RUBIN, D.B. Estimating causal effects of treatments in randomized and non-randomized studies. **Journal of Educational Psychology**, 66 (5): 688-701, 1974.

SCHULTZ, T. P.; TANSEL, A. Wage and labor supply effects of illness in Côte d'Ivoire and Ghana: Instrumental variable estimates for days disabled. **Journal of Development Economics**. v. 53, n. 2, p. 251-286, 1997.

- SCHULTZ, T. W. Investment in Human Capital. **The American Economic Review**, v. 51, n. 1, p. 1–17, 1961.
- SILVA, F. M. **Um estudo econômico da tuberculose no Brasil**. 115p. Dissertação (Mestrado em Economia do Desenvolvimento) – Faculdade de Administração, Contabilidade e Economia, PUCRS, Porto Alegre, 2011.
- SOUSA, E. A. **Efeitos da educação sobre a saúde do indivíduo: uma análise para a região nordeste do Brasil**. 2010. 82f. Dissertação (Mestrado em Economia), Faculdade de Economia, Administração e Contabilidade, UFAL, Maceió, 2010.
- STRAUSS, J.; THOMAS, D. Health, Nutrition, and Economic Development. **Journal of Economic Literature**, v. 36, n. 2, p. 766–817, 1998.
- TIGRE, R.; Sampaio, B.; MENEZES, T. The Impact of Commuting Time on Youth's School Performance. **Journal of Regional Science**, 2016.

#### Anexo A.1 – Estatísticas Descritivas – PNAD - 2008

	Variáveis	Média	Desvio-Padrão	Mínimo	Máximo	Estágio
	Renda do Trabalho (Dependente)	1,211,93	1,772,65	0	27.992	2º
	Trabalho	0,9714	0,1666	0	1	1º
Renda do Não Trabalho	Pensão	14,5992	152,2472	0	10.000	1º
	Aposentadoria	0,6679	41,8254	0	5.700	1º
	Aluguel	13,5666	168,2745	0	12.000	1º
	Doação	2,3748	57,4578	0	6.000	1º
	Juros	11,0819	76,0204	0	5.000	1º
Características Pessoais	Sexo	0,5039	0,4999	0	1	1º, 2º
	Raça	0,4564	0,4981	0	1	1º, 2º
	Idade	40,6814	10,1470	25	64	1º, 2º
	Estado Civil	0,7334	0,4421	0	1	1º, 2º
	Anos de Estudo	9,2326	4,5575	1	16	1º, 2º
Hábitos de Saúde	Atividade física	0,2885	0,4530	0	1	1º, 2º
	Fumante	0,2009	0,4007	0	1	1º, 2º
Ramo de Atividade	Agrícola	0,1093	0,3120	0	1	2º
	Transformação	0,1374	0,3443	0	1	2º
	Construção	0,0624	0,2419	0	1	2º
	Outras indústrias	0,0075	0,0865	0	1	2º
	Comércio	0,1627	0,3691	0	1	2º
	Serviços	0,0439	0,2084	0	1	2º
	Serviços auxiliares	0,0529	0,2239	0	1	2º
	Trans. e comunicação	0,0434	0,2039	0	1	2º
	Social	0,1300	0,3363	0	1	2º
	Adm. Pública	0,0697	0,2546	0	1	2º
	Outras	0,0807	0,2724	0	1	2º
Posição na Ocupação do Trabalho	Militar	0,1110	0,3142	0	1	2º
	Com carteira	0,3127	0,4636	0	1	2º
	Sem carteira	0,1472	0,3543	0	1	2º
	Doméstico	0,0995	0,2994	0	1	2º
	Conta-própria	0,2751	0,4465	0	1	2º
	Empregador	0,0542	0,2266	0	1	2º
Localização	Urbana	0,8634	0,3433	0	1	1º, 2º
	Norte	0,1298	0,3361	0	1	1º, 2º
	Nordeste	0,2854	0,4516	0	1	1º, 2º
	Sul	0,1664	0,3724	0	1	1º, 2º
	Sudeste	0,2980	0,4574	0	1	1º, 2º
	Centro-Oeste	0,1201	0,3251	0	1	1º, 2º

Fonte: Elaboração própria com base nos dados da PNAD 2008.

Nota: categorias das variáveis – Sexo (1=feminino; 0=masculino), Raça (1= brancos; 0=brancos).