



UNIVERSIDADE FEDERAL DE GOIÁS

FACULDADE DE ADMINISTRAÇÃO, CIÊNCIAS
CONTÁBEIS E ECONOMIA

Danúbia Rodrigues da Cunha

Modelos de regressão bivariada: uma aplicação em equações mincerianas de rendimento

Durante o desenvolvimento deste trabalho a autora recebeu auxílio financeiro da

CAPES

Goiânia

2018

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR
VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES
NA BIBLIOTECA DIGITAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: ☒ Dissertação ☐ Tese

2. Identificação da Tese ou Dissertação:

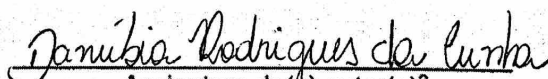
Nome completo do autor: **Danúbia Rodrigues da Cunha**

Título do trabalho: **Modelos de regressão bivariada: uma aplicação em equações mincerianas de rendimento**

3. Informações de acesso ao documento:

Concorda com a liberação total do documento ☒ SIM ☐ NÃO¹

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.


Assinatura do(a) autor(a)²

Ciente e de acordo: 
Prof. Sandro Eduardo Monsueto
Professor Adjunto SIAPE 1698093
FACE/UFG

Assinatura do(a) orientador(a)²

Data: 27 / 03 / 2018

¹ Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

² A assinatura deve ser escaneada.

Danúbia Rodrigues da Cunha

Modelos de regressão bivariada: uma aplicação em equações mincerianas de rendimento

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Economia da Faculdade de Administração, Ciências Contábeis e Economia da Universidade Federal de Goiás, como requisito parcial para obtenção de título de Mestre em Economia.

Universidade Federal de Goiás – UFG

Faculdade de Administração, Ciências Contábeis e Economia

Orientador: Prof. Dr. Sandro Eduardo Monsueto

Coorientador: Prof. Dr. Helton Saulo Bezerra dos Santos

Goiânia

2018

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Rodrigues da Cunha, Danúbia

Modelos de regressão bivariada: uma aplicação em equações mincerianas de rendimento [manuscrito] / Danúbia Rodrigues da Cunha. - 2018.

72 f.: il.

Orientador: Prof. Sandro Eduardo Monsueto; co-orientador Helton Saulo Bezerra dos Santos.

Dissertação (Mestrado) - Universidade Federal de Goiás, Faculdade de Administração, Ciências Contábeis e Ciências Econômicas (FACE), Programa de Pós-Graduação em Economia, Goiânia, 2018.

Bibliografia. Apêndice.

Inclui abreviaturas, símbolos, gráfico, tabelas, algoritmos, lista de figuras, lista de tabelas.

1. Equação de salário minceriana. 2. Modelo de regressão bivariada. I. Eduardo Monsueto, Sandro, orient. II. Título.

CDU 33



ATA DE DEFESA PÚBLICA DE DISSERTAÇÃO

No dia 08 de fevereiro de 2018, no horário de 16 horas às 17:00 horas, foi realizada, em sessão pública na sala 2110 da FACE, a defesa da dissertação *Modelos de regressão bivariada: uma aplicação em equações mincerianas de rendimento*, de autoria da Mestranda Danúbia Rodrigues da Cunha, do Programa de Pós-Graduação em Economia – PPE da Universidade Federal de Goiás. A Comissão Examinadora, constituída pelos professores Sandro Eduardo Monsueto/PPE/UFG (Orientador/membro interno), Priscila Casari/PPE/UFG (Examinadora/membro interno) e Mário Ernesto Piscoya Dias IME/UFG (Examinador/membro externo), emitiu o seguinte parecer/recomendações:

Resultado Final:

☒ Aprovado

☐ Reprovado

Recomendações:

Eu, Sandro Eduardo Monsueto, orientador da discente, lavrei a presente Ata, que segue assinada por mim e pelos demais membros da Comissão Examinadora.

Prof. Dr. Sandro Eduardo Monsueto (Orientador/Presidente/PPE/UFG)

Prof. Dr. Priscila Casari (Examinadora Interna/PPE/UFG)

Prof. Dr. Mário Ernesto Piscoya Dias (Examinadora Externo/IME/UFG)

Resumo

Nessa dissertação, modelos de regressão bivariada baseados nas distribuições bivariadas normal, t e Birnbaum-Saunders são usados para analisar dados do mercado de trabalho. Em especial, o objetivo é modelar a variável dependente da equação de rendimento minceriana de forma separada, ou seja, o rendimento-hora é modelado em duas partes, rendimento e horas trabalhadas. Os modelos de regressão bivariada são utilizados para modelar essas duas partes de forma a tentar captar a correlação entre elas e os distintos efeitos, ou seja, remuneração ou prêmio pelo esforço despendido pela mão de obra, e oferta de trabalho ou o tempo que o trabalhador disponibiliza ao mercado. Para tal, usa-se dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) para os anos de 2013, 2014 e 2015. Os parâmetros dos modelos são estimados usando o método da máxima verossimilhança. Os resultados mostram que o modelo de regressão bivariada baseada na distribuição bivariada t tem o melhor ajuste para os dados, e que a presença de correlação entre rendimento e horas trabalhadas indica que o modelo bivariado é mais adequado que o univariado.

Palavras-chave: Distribuição normal bivariada; Distribuição t bivariada; Distribuição Birnbaum-Saunders bivariada; Economia do trabalho; Equação de salário minceriana; Estimador de máxima verossimilhança; Modelo de regressão bivariada.

Abstract

In this work, bivariate regression models based on the bivariate normal, t and Birnbaum-Saunders distributions are used to analyze labor market data. In special, the objective is to model the dependent variable of the Mincerian earnings equation separately, namely, the variable hourly earnings (which is obtained by dividing gross monthly earnings by hours worked) is modeled in two parts, earnings and hours worked. The bivariate regression models are used to model these two parts in order to try to capture the correlation between them and the different effects, that is, remuneration or premium for labor effort, and the labor supply or the time that the worker offers to the market. In order to accomplish this, data from the Brazilian National Household Sample Survey (PNAD) for the years 2013, 2014 and 2015 are used. The parameters of the models are estimated using the maximum likelihood method. The results show that the bivariate regression model based on the bivariate t distribution has the best fit for the data, and that the presence of correlation between earnings and hours worked indicates that the bivariate model is more adequate than the univariate model.

Keywords: Bivariate normal distribution; Bivariate t distribution; Bivariate Birnbaum-Saunders distribution; Labor economics; Mincerian earnings equation; Maximum likelihood estimator; Bivariate regression model.

Lista de ilustrações

Figura 1	–	Histograma para rendimento e horas trabalhadas (nível e escala logarítmica).	38
Figura 2	–	Gráficos PP dos das distâncias transformadas para os modelos de regressão: univariada t do rendimento (esquerdo) e univariada t das horas trabalhadas (direito) - EC: probabilidade empírica; TC: probabilidade teórica.	40
Figura 3	–	Gráfico PP da distância transformada para o modelo de regressão bivariada t do rendimento e horas trabalhadas - EC: probabilidade empírica; TC: probabilidade teórica.	41

Lista de tabelas

Tabela 1 – Estatísticas descritivas para rendimento e horas trabalhadas (nível e escala logarítmica).	37
Tabela 2 – Resultados das critérios de informação AIC e BIC para os modelos indicados.	39
Tabela 3 – Estimativas para o modelo de regressão univariada do rendimento com a distribuição t ($\nu = 4$).	43
Tabela 4 – Estimativas para o modelo de regressão univariada das horas trabalhadas com a distribuição t ($\nu = 4$).	44
Tabela 5 – Estimativas para o modelo de regressão bivariada do rendimento e horas trabalhadas com a distribuição t ($\nu = 4$) bivariada.	48
Tabela 6 – Estimativas para o modelo de regressão univariada do rendimento-hora tradicional.	49
Tabela 7 – Estimativas para o modelo de regressão univariada do rendimento com a distribuição normal.	52
Tabela 8 – Estimativas para o modelo de regressão univariada das horas trabalhadas com a distribuição normal.	53
Tabela 9 – Estimativas para o modelo de regressão univariada do rendimento com a distribuição BS.	54
Tabela 10 – Estimativas para o modelo de regressão univariada das horas trabalhadas com a distribuição BS.	55
Tabela 11 – Estimativas para o modelo de regressão bivariada do rendimento e horas trabalhadas com a distribuição normal bivariada.	56
Tabela 12 – Estimativas para o modelo de regressão bivariada do rendimento e horas trabalhadas com a distribuição BS bivariada.	57

Lista de abreviaturas e siglas

BS	Birnbaum-Saunders
FDP	Função Densidade de Probabilidade
FDA	Função de Distribuição Acumulada
IBGE	Instituto Brasileiro de Geografia e Estatística
MQO	Mínimos Quadrados Ordinários
PP	Probabilidade-Probabilidade
PNAD	Pesquisa Nacional por Amostra de Domicílio
RAIS	Relação Anual de Informações Sociais

Sumário

1	INTRODUÇÃO	12
2	EQUAÇÃO MINCERIANA E OFERTA DE TRABALHO	15
2.1	Equação minceriana	15
2.2	Oferta de trabalho	19
2.3	Nova abordagem metodológica	20
3	DISTRIBUIÇÕES E MODELOS DE REGRESSÃO BIVARIADA	22
3.1	Distribuição normal	22
3.1.1	Distribuição normal univariada	22
3.1.2	Distribuição normal bivariada	23
3.1.3	Modelo de regressão normal bivariado	24
3.2	Distribuição t	25
3.2.1	Distribuição t univariada	25
3.2.2	Distribuição t bivariada	25
3.2.3	Modelo de regressão t bivariado	26
3.3	Distribuição Birnbaum-Saunders	27
3.3.1	Distribuição Birnbaum-Saunders univariada	27
3.3.2	Distribuição Birnbaum-Saunders bivariada	28
3.3.3	Modelo de regressão Birnbaum-Saunders bivariado	28
4	ESTRATÉGIA EMPÍRICA	30
4.1	A base de dados	30
4.2	Regressão bivariada - modelo empírico	30
4.2.1	Interpretação dos coeficientes da regressão	33
5	RESULTADOS DOS MODELOS	37
5.1	Análise exploratória dos dados	37
5.2	Resultados econométricos	38
5.2.1	Investigação do melhor ajuste	38

5.2.2	Resultados das estimativas	41
5.2.2.1	Estimativas univariadas t	42
5.2.2.2	Estimativas bivariadas t e comparação com equação minceriana tradicional	45
6	CONSIDERAÇÕES FINAIS	50
A	RESULTADOS DAS ESTIMATIVAS	52
B	CÓDIGO EM R	58
	REFERÊNCIAS	69

1 Introdução

A equação de rendimentos minceriana introduzida por Mincer (1974) é a base de uma vasta literatura empírica na área de economia. Neste modelo, a teoria moderna de capital humano¹ (BECKER, 1993; HECKMAN; TOBIAS; VYTLACIL, 2000) é utilizada para estimar os retornos de escolaridade e da experiência sobre o logaritmo do rendimento-hora² recebido por um trabalhador. A equação de rendimentos minceriana é estruturada através de um modelo de regressão linear em que a variável dependente é o rendimento-hora e as variáveis explicativas são escolaridade e experiência. Depois do modelo original proposto por Mincer (1974), uma série de novas variáveis explicativas foram incorporadas ao modelo para captar, por exemplo, efeitos de discriminação, informalidade e segregação ocupacional.

Tradicionalmente, o modelo minceriano combina dois elementos para formar a variável dependente (rendimento-hora), o qual é determinada pela divisão do rendimento pelas horas trabalhadas. Contudo, estas duas variáveis podem estar captando comportamentos distintos. Se, de um lado, a remuneração mensal representa a recompensa ou prêmio pelo esforço despendido pela mão de obra, as horas trabalhadas representam a disponibilidade de tempo que este mesmo trabalhador disponibiliza ao mercado, ou seja, o lado da oferta. Portanto, estes dois comportamentos podem ser explicados de formas distintas, algo que não é percebido pela equação tradicional de rendimento-hora. Porém, a estimativa de um modelo de regressão linear para cada componente de forma independente pode fornecer previsões erradas, dado que se espera que ambas variáveis, rendimento e horas trabalhadas, estejam também correlacionados entre si - Marchant, Leiva e Cysneiros (2016). Dessa forma, um exemplo comumente encontrado em economia e que poderia ser convertido em uma aplicação bivariada é a denominada equação minceriana de rendimento-hora.

¹ Capital humano é o conjunto de capacidades, conhecimentos, competências e atributos de personalidade que favorecem a realização de trabalho de modo a produzir valor econômico. São os atributos adquiridos por um trabalhador por meio da educação, perícia e experiência. O uso do termo na literatura econômica neoclássica foi introduzido em Mincer (1958).

² Rendimento que o trabalhador recebe por hora trabalhada, geralmente pago mensalmente.

Neste cenário, a proposta de regressão bivariada pode contribuir, uma vez que tal modelagem permite a construção de modelos com estruturas de correlação entre duas medidas tomadas dentro de comuns ou distintas unidades amostrais. Além disso, há um ganho de eficiência ao se ajustar conjuntamente os parâmetros das duas regressões em um modelo de regressão bivariada, aos invés de se ajustar duas regressões separadas. Para a investigação desse problema, faz-se necessário o conhecimento sobre distribuições na sua forma bivariada, e para o presente trabalho são usadas as seguintes distribuições: normal, t de Student (ou simplesmente t) e Birnbaum-Saunders (BS).

A distribuição normal univariada é um dos modelos mais utilizados em estatística para modelar fenômenos naturais. Essa distribuição é simétrica e foi inicialmente estudada por Abraham de Moivre, Pierre-Simon Laplace e Carl Friedrich Gauss. Em especial, o último a utilizou para descrever erros de observação na astronomia. Da mesma forma, a versão bivariada da distribuição normal também é intensamente utilizada na literatura - Johnson, Kotz e Balakrishnan (1995). Desde a origem dos seus estudos, datada da metade do século dezenove no trabalho de Laplace, Plana, Gauss e Bravais, ela tem sido utilizada como base para a teoria de mensuração de erro - Balakrishnan e Lai (2009) - e outras áreas teóricas e empíricas.

Uma alternativa simétrica à distribuição normal tanto na versão univariada como bivariada, é a distribuição t - Johnson, Kotz e Balakrishnan (1995) e Balakrishnan e Lai (2009). As distribuições t univariada e bivariada possuem caudas mais pesadas do que as distribuições normal univariada e bivariada. Tal flexibilidade é importante para acomodar observações com valores mais destoantes (outliers), o que torna a t uma alternativa de bastante interesse.

No contexto assimétrico, uma distribuição que tem recebido considerável atenção é o modelo BS, o qual foi introduzido por Birnbaum e Saunders (1969) e tem sua gênese motivada por problemas de vibração em aviões comerciais que causavam fadiga de materiais - Johnson, Kotz e Balakrishnan (1995). Recentemente, Santos-Neto et al. (2012) propuseram várias reparametrizações da distribuição BS clássica, que permitem diversas características na modelação de dados ser consideradas. Uma dessas reparametrizações, é indexada por um parâmetro de escala (que também é a média da distribuição) e um parâmetro de forma (e precisão). Nesse cenário, Saulo et al. (2016) propuseram uma distribuição BS bivariada e o seu correspondente modelo de regressão bivariada

BS, baseado na distribuição reparametrizada pela média proposta por Santos-Neto et al. (2012). Destacando que, a partir desta reparametrização, não há a necessidade de transformação da variável dependente para uma escala logarítmica, o que é uma vantagem uma vez que ela pode provocar dificuldades de interpretação.

Nesse contexto, o principal objetivo desse trabalho é utilizar modelos de regressão baseados nas distribuições bivariadas normal, t e BS para modelar a equação de rendimentos minceriana. Em geral, na modelagem econométrica dessa equação, utiliza-se como variável dependente o rendimento-hora. No entanto, nesse trabalho, a modelagem do rendimento-hora é feita de maneira separada, ou seja, utiliza-se modelagem bivariada tendo como variáveis dependentes o rendimento e as horas trabalhadas. Para fins de aplicação, são usados dados da Pesquisa Nacional por Amostragem de Domicílio (PNAD) de 2013, 2014 e 2015.

Além desta seção introdutória, o restante dessa dissertação é organizada da seguinte forma. O Capítulo 2 é dedicado a apresentar conceitos e a evolução da equação minceriana, além de fazer uma breve descrição sobre oferta de trabalho, e finaliza apresentando a nova abordagem para o tema a ser estudado. No Capítulo 3, são descritas brevemente as distribuições normal, t e BS nas suas formas univariadas e bivariadas, bem como algumas de suas propriedades. Ainda nesse capítulo, são descritos os modelos de regressão bivariada baseados nas distribuições normal, t e BS bivariadas e são abordados brevemente a estimação dos parâmetros por máxima verossimilhança. Já o Capítulo 4, visa discutir a base de dados utilizada e o modelo bivariado empregado para se estimar as equações de rendimentos. No Capítulo 5, é feita a aplicação dos modelos de regressão bivariada em equações mincerianas de rendimento, usando dados da PNAD. Finalmente, conclusões e direcionamentos para trabalhos futuros são feitos no Capítulo 6.

2 Equação minceriana e oferta de trabalho

Este capítulo destina-se a apresentar alguns conceitos e trabalhos encontrados na literatura que discutem sobre a equação de rendimento minceriana, bem como alguns métodos para sua estimação. Ainda neste capítulo, é feita uma abordagem sobre oferta de trabalho, onde é discutido a participação do indivíduo no mercado de trabalho brasileiro. Por fim, dado o contexto atual da literatura sobre equação de rendimento minceriana e oferta de trabalho, será discutida a nova abordagem metodológica introduzida nessa dissertação.

2.1 Equação minceriana

Essa seção tem como objetivo estudar o surgimento da equação de rendimento minceriana bem como trabalhos decorrentes, analisando sobretudo o caso brasileiro. Em adição, métodos de estimação para a equação de rendimento são discutidos.

Vários são os estudos desenvolvidos sobre a relação das características individuais dos trabalhadores, como educação, e seus possíveis impactos na renda - Crespo e Reis (2009), Heckman, Tobias e Vytlačil (2000) e Reis e Ramos (2011). Esses estudos fazem uma alusão ao artigo inicial de Mincer (1958), que estabelece relações teóricas que justificam a causa e efeito da experiência de trabalho e escolaridade dos indivíduos sobre os seus rendimentos.

Schultz (1973), Mincer (1958) e Becker (1993) são considerados os precursores na análise moderna do capital humano. Eles defendem a educação como forma de investimento, ou seja, essa teoria estabelece que cada trabalhador pode escolher o nível de escolaridade que deseja. E, como consequência esse investimento pode implicar em mudanças nas habilidades para o trabalho e em um aumento no nível de conhecimento para executar tarefas implicando em impactos positivos na qualidade de vida - Cunha, Cornachione-Junior e Martins (2010).

Em 1974, Mincer estabelece, pela primeira vez, uma equação para rendimentos que relaciona as características de cada indivíduo à sua renda do trabalho, integrando a teoria do investimento em capital humano em um modelo empírico. Ele examinou a questão numa perspectiva econométrica e propôs que a distribuição dos rendimentos de cada trabalhador entre as suas diferentes ocupações estaria relacionada positivamente com a quantidade de investimento feito em capital humano, o que causaria impacto na produtividade e no crescimento da economia. A partir de então, o artigo de Mincer (1974) passa a ser amplamente utilizado como base para os estudos que buscam relacionar habilidades dos trabalhadores ao nível de renda percebida pelo trabalho.

Gradativamente, novas variáveis foram sendo acrescentadas à equação minceriana, como por exemplo, variáveis binárias que são capazes de diferir características pessoais do indivíduo quanto ao gênero e raça, para captar discriminação, e também termos quadráticos para escolaridade e experiência, permitindo observar pontos críticos, a partir dos quais efeitos marginais sobre os rendimentos salariais passam a ser negativos, como em Becker (1975), Willis e Rosen (1979) e Garen (1984). A posteriori, diversos estudos empíricos, como por exemplo Cain (1986) e Sachcida, Loureiro e Mendonça (2004), utilizaram o trabalho de Mincer (1958). Eles expandem a modelagem inicial através da incorporação de elementos capazes de captar outras características individuais, tais como estado civil e região de residência para analisar efeitos da localização sobre os diferenciais de salário.

Outros autores acrescentam ainda variáveis para considerar aspectos do mercado de trabalho, como setor de atividade, o tipo de ocupação ou profissão e a modalidade de contrato existente, considerando que tais variáveis também influenciam nos rendimentos auferidos. Curi e Menezes-Filho (2004), por exemplo, mostram que a taxa de permanência no emprego sem carteira é consideravelmente inferior às taxas de permanência no emprego com carteira e no trabalho por conta própria. Já em Tannuri-Pianto e Pianto (2002), é estudado os motivos que levariam os indivíduos a atuar no setor informal da economia. Os autores analisam os efeitos de características não-observáveis, tais como habilidades específicas, sobre as escolhas dos trabalhadores em atuar no setor formal ou informal, contradizendo a hipótese de segmentação do mercado de trabalho, pela qual os diferenciais salariais estão em função das características que podem ser observáveis nos postos de trabalho.

Barros, Mendonça e Duarte (1997), dividem a população ocupada, e com o mesmo nível educacional, em 19 categorias ocupacionais para estimar o diferencial de rendimento entre trabalhadores. Os autores concluem que, quando os diferenciais são analisados dentro das ocupações, estes tendem a ser maiores nos grupos ocupacionais de maior nível salarial e, quando uma desagregação por sexo é empregada, é possível notar que a estrutura das ocupações se faz mais importante para as mulheres, que explica 29% da desigualdade, enquanto que para os homens explica 12%.

De acordo com Oliveira (1998), uma das dimensões da desigualdade salarial por gênero ou raça no mercado de trabalho se revela no ingresso diferenciado à determinadas ocupações. O estudo da autora, que se baseia nos dados da PNAD para o ano de 1995 e utiliza o índice de dissimilaridade de Duncan e Duncan, evidencia um elevado nível de segregação ocupacional por sexo e raça no mercado de trabalho brasileiro, principalmente para as mulheres e indivíduos não-brancos que estão inseridos em ocupações menos favorecidas. E o estudo realizado por Matos e Machado (2006), com base nos dados da PNAD entre 1987 e 2001, destaca que as mulheres negras possuem a pior situação em termos de salário e ocupação. Se comparadas com as mulheres brancas, as negras apresentam um menor nível de escolaridade e estão mais sujeitas a discriminação e segregação ocupacional. Isso pode ser observado também em Vaz e Hoffmann (2011) que identificam a formação de “guetos” ocupacionais com elevada concentração de mulheres, a partir dos dados da RAIS para os anos de 1995 e 2008.

No que tange aos métodos de estimação, desde Mincer (1974) a literatura tem utilizado uma série de modelos econométricos, desde o tradicional mínimos quadrados ordinários (MQO) e sua versão com variáveis instrumentais, ou versões mais robustas, como regressão quantílica, seleção amostral e procedimentos baseados no método de estimação da máxima verossimilhança - Chatterjee e Price (1991), Stapleton (1995), Heckman (1974), Heckman (1976), Heckman (1979), Garen (1984), Buchinsky (2001).

No Brasil, o uso de regressões de rendimento tem início com o estudo de Langoni (1973) que estima seu modelo por meio de regressões log-lineares tendo como variáveis explicativas a educação, idade, sexo, atividade e região que o trabalhador reside. Langoni considerava que a educação tinha grande relevância como variável explicativa da renda. Senna (1976) e Branco (1979), também adeptos da teoria do capital humano, analisam os

dados da Lei dos Dois Terços¹, de 1970, e avaliam os efeitos da educação e da experiência sobre os rendimentos do trabalho. Branco (1979) utiliza regressões de rendimentos sobre a escolaridade e a experiência no emprego para obter as estimativas para o retorno do investimento em capital humano, obtendo como resultado um aumento da taxa de retorno à educação entre os anos da amostra, além de uma variação dos retornos entre os setores de atividade.

A maior disponibilidade de microdados e aumento da capacidade e facilidade computacional ajudou a promover uma expansão dos estudos brasileiros sobre equações de rendimento a partir dos anos 2000. Estimções por regressões quantílicas são realizadas, por exemplo, em Maciel, Campêlo e Raposo (2001) que utilizam equações de rendimento para verificar mudanças ocorridas nos retornos à educação no rendimento das mulheres ocupadas utilizando microdados da PNAD de 1992 e 1999. Em Monsueto (2003), ela é utilizada para captar efeitos diferenciados dos fatores em função da posição do trabalhador na distribuição de renda.

Um ponto em comum com a maioria dos trabalhos sobre a determinação do rendimento usando a equação minceriana, tanto na literatura nacional como internacional, é o uso do rendimento-hora como variável dependente. Esta, em geral, é obtida por meio da simples divisão do rendimento mensal pelas horas trabalhadas no período. Tal abordagem implica na aglutinação em uma única variável (rendimento-hora) de dois componentes distintos, e que a princípio, poderiam ser modelados separadamente.

Em geral, as variáveis rendimento e horas trabalhadas podem estar captando fenômenos e efeitos distintos sobre o trabalhador. Se, de um lado, a remuneração ou o salário representa o prêmio ou compensação pelo esforço da mão de obra, as horas de trabalho captam o tempo disponibilizado pelo trabalhador ao mercado, ou seja, o lado da oferta de trabalho. Neste sentido, elementos que modificam os salários podem andar de forma distinta sobre a oferta de trabalho, algo que não é captado pela equação de rendimento minceriana tradicional bem como por todas as equações que usam o rendimento-hora como variável dependente.

A seguir, faz-se uma breve discussão sobre oferta de trabalho, em particular, a participação do indivíduo no mercado de trabalho brasileiro.

¹ Decreto nº. 19.482, de 12-12-1930, onde as empresas deveriam ter em seus quadros de empregados dois terços de trabalhadores nacionais.

2.2 Oferta de trabalho

Quando se pensa na decisão de participação no mercado de trabalho, é sabido que ela está intimamente ligada à oferta de trabalho e esta seção tem como objetivo apresentar algumas evidências empíricas sobre o lado da oferta com trabalhos que utilizam microdados, buscando uma melhor compreensão de como se dividem as decisões de participação na força de trabalho entre os indivíduos cujos recursos se agregam no contexto do domicílio.

No Brasil, alguns trabalhos tratam dos impactos na decisão de participação no mercado de trabalho. Sedlacek e Santos (1991), utilizando dados da PNAD de 1984, analisam a relação entre o rendimento do marido e participação das mulheres cônjuges. Os autores concluem que, quanto maior a renda do marido, maior seria o salário de reserva e menor a probabilidade de as mulheres trabalharem, e que quanto mais jovens e mais numerosos os filhos da família, menor é a probabilidade delas entrarem no mercado de trabalho. Mas, quanto maior a educação dessas mulheres, maior a probabilidade de trabalharem.

Já a partir dos trabalhos de Barros, Ramos e Santos (1995), Barros, Mendonça e Duarte (1997) e Junior, Menezes-Filho e Ferreira (2005), observa-se que as mulheres são menos propensas à participação no mercado de trabalho e à aquisição de um emprego do que os homens. Além disso, nota-se que as chances de contratação aumentam com a elevação dos anos de estudo e com uma faixa de idade até os 40 anos. Entretanto, característica como residência em área urbana e/ou metropolitanas afeta negativamente essas chances, resultado que ratifica a existência de maior competição por emprego e o maior nível de desemprego em tais regiões. Tais resultados revelam a importância do investimento em capital humano e são consistentes com aqueles verificados em vários estudos sobre mercado de trabalho e diferenciais salariais no Brasil.

Com base nas duas revisões apresentadas, é possível considerar que os determinantes da oferta de trabalho e do rendimento podem não ser os mesmos, tanto em termos de quantidade, sinais e magnitudes. Isto é, horas trabalhadas e rendimento mensal possuem fatores determinantes distintos, algo não captado de maneira direta pelas tradicionais equações mincerianas. Por exemplo, o estoque de capital humano, captado pela educação formal ou pela experiência tende a aumentar a remuneração dos trabalha-

dores, mas pode ter efeito contrário sobre a disposição a ofertar horas ao mercado de trabalho. Surge, assim, uma lacuna na literatura a qual esse trabalho objetiva preencher utilizando uma abordagem bivariada em contraposição à clássica versão de Mincer para rendimento-hora.

Na seção a seguir, é discutida essa nova abordagem metodológica proposta nessa dissertação a fim de explicar conjuntamente rendimento e oferta de trabalho.

2.3 Nova abordagem metodológica

Nas seções anteriores foram discutidos, entre outros elementos, dois pontos principais: a) os determinantes do rendimento através do estudo da equação minceriana; e (b) os determinantes da oferta de trabalho dada à decisão de participação dos indivíduos no mercado de trabalho. Nesse dissertação, uma abordagem diferente é dada a equação de rendimento minceriana. Se, por um lado, boa parte dos trabalhos toma a variável dependente como rendimento-hora em logaritmo, por outro, tal abordagem pode mascarar/ocultar efeitos de remuneração e oferta de trabalho. Assim, esse trabalho utiliza o arcabouço de regressão bivariada, ou seja, regressão para um vetor bivariado de variáveis dependentes, para modelar de forma separada rendimento e horas trabalhadas. Aqui, tais variáveis dependentes são relacionadas diretamente aos seus determinantes através de uma estrutura de regressão, e é permitido uma correlação entre tais variáveis.

A partir disso, é possível elencar as vantagens associadas à abordagem de regressão bivariada como, por exemplo, a possibilidade de modelos com estrutura de correlação entre as variáveis dependentes, ou seja, a possibilidade de mensuração da correlação existente entre variáveis dependentes. Como mencionado no Capítulo 1, se existir correlação entre as variáveis dependentes, a estimação isoladamente de dois modelos de regressão univariada para rendimento e horas trabalhadas pode fornecer previsões incorretas. Uma outra vantagem, é a captação de efeitos de recompensa ou prêmio pelo esforço despendido pela mão de obra e disponibilidade de tempo que este mesmo trabalhador disponibiliza ao mercado, ou seja, aqui a modelagem do rendimento e horas trabalhadas no arcabouço bivariado permite o estudo dos efeitos das variáveis explicativas nessas variáveis. E, por fim, há a vantagem na possibilidade de assumir diferentes distribuições bivariadas (diferentes distribuições flexíveis podem ser assumidas) para modelagem do

rendimento e horas trabalhadas. Assim como em Heckman (1976), as estimativas dos parâmetros dos modelos são obtidos pelo método da máxima verossimilhança, o que pode ser visto como mais uma vantagem, uma vez que esse é um método de estimação eficiente - Mittelhammer, Judge e Miller (2000).

Como mencionado, é possível assumir diferentes distribuições bivariadas para a modelagem das variáveis dependentes. Em particular, são utilizadas e comparadas três distribuições, a saber, normal, t e BS. É esperado que o uso da modelagem proposta seja uma alternativa mais adequada do que aquela tradicionalmente utilizada na literatura (equação minceriana de rendimento-hora). A seguir, no Capítulo 3, são detalhadas tais distribuições bem como os modelos de regressão bivariado correspondentes. Ainda nesse capítulo, discussões sobre estimações dos parâmetros em cada modelo são apresentados.

3 Distribuições e modelos de regressão bivariada

No capítulo anterior foi analisado a importância de se estimar separado renda e horas trabalhadas. Contudo, também é salientado que ambos podem ser correlacionados e que uma equação isolada pode levar a previsões incorretas. Desta forma, é proposto uma abordagem bivariada. O presente capítulo mostra a parte técnica desta abordagem.

Nesse capítulo é feita a descrição das distribuições normal, t e BS nas suas correspondentes versões univariadas e bivariadas. As distribuições bivariadas são utilizadas para obtenção dos correspondentes modelos de regressão bivariada normal (Seção 3.1.3), t (Seção 3.2.3) e BS (Seção 3.3.3) - Johnson, Kotz e Balakrishnan (1995), Balakrishnan e Lai (2009) e Saulo et al. (2016).

3.1 Distribuição normal

3.1.1 Distribuição normal univariada

Uma variável aleatória Y segue uma distribuição normal com média $\mu \in \mathbb{R}$ e desvio-padrão $\sigma^2 > 0$, denotada por $Y \sim N(\mu, \sigma^2)$, se a função densidade de probabilidade (FDP) de Y é dada por

$$f(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{y - \mu}{\sigma} \right)^2 \right\}, \quad -\infty < y < \infty. \quad (3.1)$$

A distribuição normal padrão, denotada por $Z \sim N(0, 1)$, é obtida quando $\mu = 0$ e $\sigma = 1$, ou seja, quando temos uma normal com média zero e desvio-padrão unitário com correspondente FDP dada por

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} z^2 \right\}, \quad -\infty < z < \infty. \quad (3.2)$$

Note que é possível obter a variável aleatória Y por uma transformação linear de Z , ou seja, $Y = \mu + \sigma Z$. A função de distribuição acumulada (FDA) associadas a Y e Z são

dadas respectivamente por

$$F(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^y \exp \left\{ -\frac{1}{2} \left(\frac{w - \mu}{\sigma} \right)^2 \right\} dw \quad (3.3)$$

e

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp \left\{ -\frac{1}{2} w^2 \right\} dw. \quad (3.4)$$

3.1.2 Distribuição normal bivariada

Seja $\mathbf{Y} = (Y_1, Y_2)^\top$ um vetor aleatório bivariado seguindo uma distribuição normal bivariada com médias μ_1 e μ_2 , e desvios-padrões σ_1 e σ_2 . Em adição a esses quatro parâmetros há um coeficiente de correlação ρ entre Y_1 e Y_2 definido por $-1 < \rho < 1$. Assim, denotando $\mathbf{Y} \sim \text{NBiv}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Então, a FDP conjunta de Y_1 e Y_2 pode ser escrita como

$$f(y_1, y_2; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(y_1-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} \right] \right\}. \quad (3.5)$$

A FDP conjunta do vetor aleatório $\mathbf{Z} = (Z_1, Z_2)^\top$ seguindo uma distribuição normal bivariada padrão (médias zero e variâncias um) é dada por

$$\phi_2(z_1, z_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (z_1^2 + z_2^2 - 2\rho z_1 z_2) \right\}. \quad (3.6)$$

A correspondente FDA conjunta associada a (3.5) é dada por

$$\Phi_2(z_1, z_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} \exp \left\{ -\frac{1}{2(1-\rho^2)} (u^2 + v^2 - 2\rho uv) \right\} du dv. \quad (3.7)$$

Quando $\rho = 0$, ou seja, quando as variáveis normais são não correlacionadas, (3.5) pode ser expressada como o produto de duas FDAs normais.

Na discussão sobre a distribuição bivariada normal, outro parâmetro de interesse é a covariância entre Y_1 e Y_2 , tal medida é definida pela esperança ou valor esperado de $(Y_1 - \mu_1)(Y_2 - \mu_2)$. Em termos dos parâmetros da distribuição normal bivariada, tem-se que a covariância (Cov) entre Y_1 e Y_2 é dada por

$$\text{Cov}[Y_1, Y_2] = \rho\sigma_1\sigma_2.$$

Note que é possível expressar as variâncias e covariâncias em uma matrix simétrica, ou seja,

$$\begin{pmatrix} \text{Var}[Y_1] & \text{Cov}[Y_1, Y_2] \\ \text{Cov}[Y_1, Y_2] & \text{Var}[Y_2] \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

3.1.3 Modelo de regressão normal bivariado

Considere $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ tal que $\mathbf{Y}_i = (Y_{1i}, Y_{2i})^\top$ siga uma distribuição normal bivariada, ou seja, $\mathbf{Y}_i \sim \text{NBiv}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Considere que há r e s covariáveis, digamos $\mathbf{x}_i^{(1)} = (x_{i1}^{(1)}, x_{i2}^{(1)}, \dots, x_{ir}^{(1)})^\top$ e $\mathbf{x}_i^{(2)} = (x_{i1}^{(2)}, x_{i2}^{(2)}, \dots, x_{is}^{(2)})^\top$, associadas com Y_{1i} e Y_{2i} , respectivamente, tal que

$$E[Y_{1i} | \mathbf{x}_i^{(1)}] = \mu_{1i} = \mathbf{x}_i^{(1)\top} \boldsymbol{\beta}_1 = \beta_{11}x_{i1}^{(1)} + \beta_{12}x_{i2}^{(1)} + \dots + \beta_{1l}x_{il}^{(1)}, \quad i = 1, \dots, n, \quad (3.8)$$

$$E[Y_{2i} | \mathbf{x}_i^{(2)}] = \mu_{2i} = \mathbf{x}_i^{(2)\top} \boldsymbol{\beta}_2 = \beta_{21}x_{i1}^{(2)} + \beta_{22}x_{i2}^{(2)} + \dots + \beta_{2l}x_{il}^{(2)}, \quad i = 1, \dots, n, \quad (3.9)$$

em que $\boldsymbol{\beta}_k = (\beta_{k1}, \beta_{k2}, \dots, \beta_{kl})$ é um vetor de l parâmetros desconhecidos, e $\mathbf{x}_i^{(k)}$ é a i -ésima linha da matriz $\mathbf{X}^{(k)}$, cuja dimensão é $n \times l$, para $k = 1, 2$ e $l = r, s$. Desse modo, temos o seguinte modelo de regressão normal bivariado

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i^{(1)\top} \boldsymbol{\beta}_1 \\ \mathbf{x}_i^{(2)\top} \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \quad i = 1, \dots, n, \quad (3.10)$$

em que $(\epsilon_{1i}, \epsilon_{2i}) \sim \text{NBiv}(0, 0, \sigma_1, \sigma_2, \rho)$, e eles independentemente distribuídos.

Para estimar os parâmetros desconhecidos $\sigma_1, \sigma_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ e ρ baseado em uma amostra aleatória de tamanho n , ou seja, $\{(y_{1i}, y_{2i}, \mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}); i = 1, \dots, n\}$, o método da máxima verossimilhança é usado. As funções de verossimilhança e log-verossimilhança da amostra observada podem ser escritas respectivamente como

$$L = \prod_{i=1}^n f(y_{1i}, y_{2i}; \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \rho), \quad (3.11)$$

$$\ell = \sum_{i=1}^n \log(f(y_{1i}, y_{2i}; \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \rho)), \quad (3.12)$$

em que f é FDP conjunta da distribuição normal bivariada. As estimativas dos parâmetros do modelo devem ser obtidas maximizando a função de log-verossimilhança (3.12). Isso é feito resolvendo um processo iterativo não-linear de otimização, em particular pode-se usar o método quase-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS). O método BFGS está implementado no software R (<http://cran.r-project.org>), por meio das funções `optim` and `optimx`.

3.2 Distribuição t

3.2.1 Distribuição t univariada

Considere uma variável aleatória U seguindo uma distribuição t com graus de liberdade ν , denotada por $U \sim t(\mu, \sigma, \nu)$. Então a FDP de U é dada por

$$f(u; \mu, \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{u^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < u < \infty. \quad (3.13)$$

Note que a distribuição t converge para uma distribuição normal quando $\nu \rightarrow \infty$, ou seja,

$$\lim_{\nu \rightarrow \infty} f(u; \mu, \sigma, \nu) = \frac{1}{\sqrt{2\pi}} \exp\{-u^2/2\}. \quad (3.14)$$

A FDA associadas a U é dada por

$$F(u; \mu, \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^u \left(1 + \frac{w^2}{\nu}\right)^{-(\nu+1)/2} dw. \quad (3.15)$$

3.2.2 Distribuição t bivariada

Seja $\mathbf{U} = (U_1, U_2)^\top$ um vetor aleatório bivariado seguindo uma distribuição t bivariada com parâmetros de localização μ_1 e μ_2 , escala σ_1 e σ_2 , graus de liberdade ν , e coeficiente de correlação $-1 < \rho < 1$ entre U_1 e U_2 , denotado por $\mathbf{U} \sim t\text{Biv}(\mu_1, \mu_2, \sigma_1, \sigma_2, \nu, \rho)$. Então, a FDP conjunta de U_1 e U_2 é dada por

$$f(u_1, u_2; \mu_1, \mu_2, \sigma_1, \sigma_2, \nu, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \left[1 + \frac{1}{\nu(1-\rho^2)} \left(\frac{(u_1-\mu_1)^2}{\sigma_1^2} + \frac{(u_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(u_1-\mu_1)(u_2-\mu_2)}{\sigma_1\sigma_2} \right) \right]^{-(\nu+2)/2}. \quad (3.16)$$

A FDP conjunta do vetor aleatório $\mathbf{U} = (U_1, U_2)^\top$ seguindo uma distribuição t bivariada padrão, o qual é a versão bivariada de (3.13), é dada por

$$f(u_1, u_2, \nu, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \left[1 + \frac{1}{\nu(1-\rho^2)} (u_1^2 + u_2^2 - 2\rho u_1 u_2) \right]^{-(\nu+2)/2}.$$

A correspondente FDA conjunta associada a (3.17) é dada por

$$F(u_1, u_2; \nu, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{u_1} \int_{-\infty}^{u_2} \left[1 + \frac{1}{\nu(1-\rho^2)} (u^2 + v^2 - 2\rho uv) \right]^{-(\nu+2)/2} dv du \quad (3.17)$$

3.2.3 Modelo de regressão t bivariado

Considere $\mathbf{U}_1, \dots, \mathbf{U}_n$ tal que $\mathbf{U}_i = (U_{1i}, U_{2i})^\top$ segue uma distribuição t bivariada, ou seja, $\mathbf{U}_i \sim t\text{Biv}(\mu_1, \mu_2, \sigma_1, \sigma_2, \nu, \rho)$ com FDP (3.16). Assuma que há r e s covariáveis, digamos $\mathbf{x}_i^{(1)} = (x_{i1}^{(1)}, x_{i2}^{(1)}, \dots, x_{ir}^{(1)})^\top$ e $\mathbf{x}_i^{(2)} = (x_{i1}^{(2)}, x_{i2}^{(2)}, \dots, x_{is}^{(2)})^\top$, associadas com U_{1i} e U_{2i} , respectivamente, tal que

$$E[U_{1i} | \mathbf{x}_i^{(1)} = \mathbf{x}_i^{(1)}] = \mu_{1i} = \mathbf{x}_i^{(1)\top} \boldsymbol{\beta}_1 = \beta_{11}x_{i1}^{(1)} + \beta_{12}x_{i2}^{(1)} + \dots + \beta_{1l}x_{il}^{(1)}, i = 1, \dots, n, \quad (3.18)$$

$$E[U_{2i} | \mathbf{x}_i^{(2)} = \mathbf{x}_i^{(2)}] = \mu_{2i} = \mathbf{x}_i^{(2)\top} \boldsymbol{\beta}_2 = \beta_{21}x_{i1}^{(2)} + \beta_{22}x_{i2}^{(2)} + \dots + \beta_{2l}x_{il}^{(2)}, i = 1, \dots, n, \quad (3.19)$$

em que $\boldsymbol{\beta}_k = (\beta_{k1}, \beta_{k2}, \dots, \beta_{kl})$ é um vetor de l parâmetros desconhecidos, e $\mathbf{x}_i^{(k)}$ é a i -ésima linha da matriz $\mathbf{X}^{(k)}$, cuja dimensão é $n \times l$, para $k = 1, 2$ e $l = r, s$. Então, tem-se o seguinte modelo de regressão t bivariado

$$\begin{pmatrix} U_{1i} \\ U_{2i} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i^{(1)\top} \boldsymbol{\beta}_1 \\ \mathbf{x}_i^{(2)\top} \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \quad i = 1, \dots, n, \quad (3.20)$$

em que $(\epsilon_{1i}, \epsilon_{2i}) \sim t\text{Biv}(0, 0, \sigma_1, \sigma_2, \nu, \rho)$ são independentemente distribuídos.

Os parâmetros do modelo são estimados similarmente ao caso normal bivariado, ou seja, dado as funções de verossimilhança e log-verossimilhança,

$$L = \prod_{i=1}^n f(u_{1i}, u_{2i}; \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \nu, \rho), \quad (3.21)$$

$$\ell = \sum_{i=1}^n \log(f(u_{1i}, u_{2i}; \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \nu, \rho)), \quad (3.22)$$

respectivamente, em que f é FDP conjunta da distribuição t bivariada, as estimativas dos parâmetros do modelo, σ_1 , σ_2 , $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ e ρ , são obtidas maximizando a função de log-verossimilhança resolvendo um processo iterativo não-linear de otimização, em particular, o método quase-Newton BFGS. O parâmetro ν é estimado de acordo com Barros, Paula e Leiva (2008). Usa-se a log-verossimilhança perfilada e os seguintes passos:

- 1) Seja $\nu_k = k$ e para cada k , $k = 1, \dots, 20$, compute as estimativas dos parâmetros $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \sigma_1, \sigma_2, \nu, \rho)^\top$ usando o método da máxima verossimilhança. Compute também a função de log-verossimilhança;
- 2) A estimativa final de ν é aquela que maximiza a função de log-verossimilhança e as estimativas associadas de $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \sigma_1, \sigma_2, \nu, \rho)^\top$ são as finais.

3.3 Distribuição Birnbaum-Saunders

3.3.1 Distribuição Birnbaum-Saunders univariada

A distribuição BS está relacionada com a distribuição normal por meio da representação estocástica

$$T = \frac{\beta}{4} \left[\alpha Z + \sqrt{(\alpha Z)^2 + 4} \right]^2,$$

em que $\alpha > 0$ e $\beta > 0$ são parâmetros de forma e escala, respectivamente, $Z \sim N(0, 1)$ e T tem distribuição BS com notação $T \sim \text{BS}(\alpha, \beta)$. A FDP de T é dada por

$$f(t; \alpha, \beta) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2\alpha^2} \left[\frac{t}{\beta} + \frac{\beta}{t} - 2 \right] \right) \frac{t^{-3/2}(t + \beta)}{2\alpha\sqrt{\beta}}, \quad t > 0. \quad (3.23)$$

A média e variância de T são dadas por $E[T] = \beta[1 + \alpha^2/2]$ e $\text{Var}[T] = [\alpha\beta]^2[1 + 5\alpha^2/4]$, respectivamente. O parâmetro de escala β é também a mediana da distribuição. A distribuição BS possui a propriedade recíproca, i.e., $1/T$ tem a mesma distribuição de T com parâmetro β substituído por $1/\beta$, $1/T \sim \text{BS}(\alpha, 1/\beta)$.

Nesse trabalho será utilizada a versão bivariada da distribuição BS baseada na reparametrização introduzida por Santos-Neto et al. (2012). Essa reparametrização é útil porque podemos fazer modelam sem a necessidade de aplicação do log nas variáveis respostas, como acontece no caso da parametrização original; ver Leiva et al. (2014). Aqui temos $\alpha = \sqrt{2/\delta}$ e $\beta = \delta\mu/(\delta + 1)$, de maneira que $\delta = 2/\alpha^2$ e $\mu = \beta[1 + \alpha^2/2]$, em que $\delta > 0$ é um parâmetro de forma e precisão e $\mu > 0$ é um parâmetro de escala e também é média da distribuição. A FDP de $T \sim \text{BS}(\mu, \delta)$ é dada por

$$f(t; \mu, \delta) = \frac{\exp(\delta/2)\sqrt{\delta+1}}{4\sqrt{\pi\mu}t^{3/2}} \left[t + \frac{\delta\mu}{\delta+1} \right] \exp \left(-\frac{\delta}{4} \left[\frac{t\{\delta+1\}}{\delta\mu} + \frac{\delta\mu}{t\{\delta+1\}} \right] \right), \quad t > 0. \quad (3.24)$$

A média e variância de T baseada na reparametrização proposta por Santos-Neto et al. (2012) são dadas por $E[T] = \mu$ e $\text{Var}[T] = \mu^2/\phi$, respectivamente, em que $\phi = \frac{[\delta+1]^2}{[2\delta+5]}$. Como mencionado, δ pode ser interpretado como um parâmetro de precisão, ou seja, para valores fixos de μ , quando $\delta \rightarrow \infty$, a variância de T tende a zero. Também, para valores fixos de μ , se $\delta \rightarrow 0$, então $\text{Var}[T] \rightarrow 5\mu^2$. A distribuição BS reparametrizada em (3.24) preserva a propriedade recíproca, isto é, se $T \sim \text{BS}(\mu, \delta)$, então $1/T \sim \text{BS}([\delta+1]^2/\mu\delta^2, \delta)$.

A seguir é apresentado a distribuição BS bivariada proposta por Saulo et al. (2016) e o respectivo modelo de regressão bivariado.

3.3.2 Distribuição Birnbaum-Saunders bivariada

Seja $\mathbf{T} = (T_1, T_2)^\top$ um vetor aleatório bivariado seguindo uma distribuição BS bivariada parameterizada pelas médias com parâmetros $\mu_1, \mu_2, \delta_1, \delta_2$ e ρ . Então, a FDA conjunta de T_1 e T_2 pode ser escritas como (SAULO et al., 2016)

$$F(t_1, t_2; \mu_1, \mu_2, \delta_1, \delta_2, \rho) = \Phi_2 \left(\sqrt{\frac{\delta_1}{2}} \left[\sqrt{\frac{(\delta_1+1)t_1}{\delta_1\mu_1}} - \sqrt{\frac{\delta_1\mu_1}{(\delta_1+1)t_1}} \right], \right. \\ \left. \sqrt{\frac{\delta_2}{2}} \left[\sqrt{\frac{(\delta_2+1)t_2}{\delta_2\mu_2}} - \sqrt{\frac{\delta_2\mu_2}{(\delta_2+1)t_2}} \right]; \rho \right), t_1, t_2 > 0, \quad (3.25)$$

em que $\mu_1 > 0, \delta_1 > 0, \mu_2 > 0, \delta_2 > 0, -1 < \rho < 1, \Phi_2$ é a FDA da distribuição bivariada normal padrão dada em (3.7). Assim, a FDP conjunta associada a (3.25) é dada por

$$f(t_1, t_2; \mu_1, \mu_2, \delta_1, \delta_2, \rho) = \phi_2 \left(\sqrt{\frac{\delta_1}{2}} \left[\sqrt{\frac{(\delta_1+1)t_1}{\delta_1\mu_1}} - \sqrt{\frac{\delta_1\mu_1}{(\delta_1+1)t_1}} \right], \right. \\ \left. \sqrt{\frac{\delta_2}{2}} \left[\sqrt{\frac{(\delta_2+1)t_2}{\delta_2\mu_2}} - \sqrt{\frac{\delta_2\mu_2}{(\delta_2+1)t_2}} \right]; \rho \right) \frac{\sqrt{\delta_1}}{2\sqrt{2}t_1} \left[\sqrt{\frac{(\delta_1+1)t_1}{\delta_1\mu_1}} + \sqrt{\frac{\delta_1\mu_1}{(\delta_1+1)t_1}} \right] \\ \frac{\sqrt{\delta_2}}{2\sqrt{2}t_2} \left[\sqrt{\frac{(\delta_2+1)t_2}{\delta_2\mu_2}} + \sqrt{\frac{\delta_2\mu_2}{(\delta_2+1)t_2}} \right], \quad (3.26)$$

em que ϕ_2 é a FDP de uma distribuição bivariada normal dada por em (3.6). Denota-se a distribuição BS bivariada com FDP (3.26) por $\text{BSBiv}(\mu_1, \mu_2, \delta_1, \delta_2, \rho)$.

3.3.3 Modelo de regressão Birnbaum-Saunders bivariado

Considere $\mathbf{T}_1, \dots, \mathbf{T}_n$ tal que $\mathbf{T}_i = (T_{1i}, T_{2i})^\top$ segue uma distribuição BS bivariada, isto é, $\mathbf{T}_i \sim \text{BSBiv}(\mu_1, \mu_2, \delta_1, \delta_2, \rho)$. Assuma que há r e s covariáveis, $\mathbf{x}_i^{(1)} = (x_{i1}^{(1)}, x_{i2}^{(1)}, \dots, x_{ir}^{(1)})^\top$ e $\mathbf{x}_i^{(2)} = (x_{i1}^{(2)}, x_{i2}^{(2)}, \dots, x_{is}^{(2)})^\top$, associadas com T_{1i} e T_{2i} , respectivamente. Então, de (3.25), temos a FDA conjunta (Saulo et al. (2016))

$$F(t_{1i}, t_{2i}; \mu_{1i}, \mu_{2i}, \delta_1, \delta_2, \rho) = \Phi_2 \left(\sqrt{\frac{\delta_1}{2}} \left[\sqrt{\frac{(\delta_1+1)t_{1i}}{\delta_1\mu_{1i}^i}} - \sqrt{\frac{\delta_1\mu_{1i}^i}{(\delta_1+1)t_{1i}}} \right], \right. \\ \left. \sqrt{\frac{\delta_2}{2}} \left[\sqrt{\frac{(\delta_2+1)t_{2i}}{\delta_2\mu_{2i}^i}} - \sqrt{\frac{\delta_2\mu_{2i}^i}{(\delta_2+1)t_{2i}}} \right]; \rho \right), t_{1i}, t_{2i} > 0 \quad i = 1, \dots, n, \quad (3.27)$$

em que

$$E[T_{1i}|\mathbf{X}_i^{(1)} = \mathbf{x}_i^{(1)}] = \mu_{1i} = \exp(\mathbf{x}_i^{(1)\top} \boldsymbol{\beta}_1) = \exp(\beta_{11}x_{i1}^{(1)} + \beta_{12}x_{i2}^{(1)} + \cdots + \beta_{1l}x_{il}^{(1)}), i = 1, \dots, n, \quad (3.28)$$

$$E[T_{2i}|\mathbf{X}_i^{(2)} = \mathbf{x}_i^{(2)}] = \mu_{2i} = \exp(\mathbf{x}_i^{(2)\top} \boldsymbol{\beta}_2) = \exp(\beta_{21}x_{i1}^{(2)} + \beta_{22}x_{i2}^{(2)} + \cdots + \beta_{2l}x_{il}^{(2)}), i = 1, \dots, n, \quad (3.29)$$

em que $\boldsymbol{\beta}_k = (\beta_{k1}, \beta_{k2}, \dots, \beta_{kl})$ é um vetor de l parâmetros desconhecidos, e $\mathbf{x}_i^{(k)}$ é a i -ésima linha da matriz $\mathbf{X}^{(k)}$, cuja dimensão é $n \times l$, para $k = 1, 2$ e $l = r, s$. Aqui, diferentemente do modelo de regressão BS baseado na parametrização clássica (RIECK; NEDELMAN, 1991), não é preciso fazer a transformação logarítmica, ou seja, os dados para a resposta são trabalhados em sua escala original.

Para estimar os parâmetros, assim como no caso bivariado normal, usa-se o método da máxima verossimilhança. Seja uma amostra aleatória de tamanho n , ou seja, $\{(t_{1i}, t_{2i}, \mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}); i = 1, \dots, n\}$, então as funções de verossimilhança e log-verossimilhança da amostra observada são dadas respectivamente por

$$L = \prod_{i=1}^n f(t_{1i}, t_{2i}; \mu_{1i}, \mu_{2i}, \delta_1, \delta_2, \rho), \quad (3.30)$$

$$\ell = \sum_{i=1}^n \log(f(t_{1i}, t_{2i}; \mu_{1i}, \mu_{2i}, \delta_1, \delta_2, \rho)), \quad (3.31)$$

em que f é FDP conjunta da distribuição BS bivariada. As estimativas dos parâmetros β_1 , β_2 , δ_1 , δ_2 e ρ são obtidas maximizando a função de log-verossimilhança (3.31) usando um processo iterativo não-linear de otimização, nesse caso, o método quase-Newton BFGS.

Esse capítulo apresentou os modelos de regressão bivariada normal, t e BS, que serão utilizados nos Capítulos 4 e 5 para formulação da modelagem do rendimento e das horas trabalhadas, bem como para estimação e avaliação dos modelos. Em particular, no Capítulo 4 será feita uma descrição dos dados e a formulação e interpretação dos modelos de regressão bivariada segundo as distribuições consideradas nesse capítulo. No Capítulo 5, será apresentada a análise exploratória dos dados, a escolha do melhor modelo usando os critérios de informação e a verificação gráfica do ajuste. O Capítulo 5 apresentará ainda a interpretação dos resultados das estimativas dos modelos descritos nesse capítulo.

4 Estratégia Empírica

Essa capítulo visa discutir a base de dados utilizada e questões metodológicas associadas aos modelos de regressão bivariada considerados. No caso dos modelos, as variáveis dependentes e independentes são descritas e a interpretação dos coeficientes da regressão é apresentada.

4.1 A base de dados

Para aplicar a modelagem bivariada entre rendimento e horas trabalhadas são utilizados dados da PNAD para os anos de 2013, 2014 e 2015, elaborado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), o qual reporta características demográficas e socioeconômicas da população brasileira anualmente, recolhendo informações como, sexo, idade, educação, trabalho e rendimento, além de características do domicílio. Nesta dissertação trabalha-se com uma amostra composta por indivíduos ocupados com idade entre 18 e 65 anos, com informações completas de rendimento e horas de trabalho. No total, a amostra contém 167.271 observações. Os dados são baseados nas dez regiões metropolitanas (RMs) brasileiras (Pará, Ceará, Pernambuco, Bahia, Minas Gerais, Rio de Janeiro, Paraná, Rio Grande do Sul, Distrito Federal e São Paulo). Os valores nominais do rendimento são deflacionados pelo Índice Nacional de Preços ao Consumidor (INPC) para o ano de 2015.

4.2 Regressão bivariada - modelo empírico

A equação de rendimento minceriana é tipicamente escrita pela seguinte regressão univariada

$$\log(\mathbf{w}) = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad (4.1)$$

em que $\log(\mathbf{w})$ é um vetor contendo o log do rendimento-hora (variável dependente) recebidos pelos trabalhadores, $\boldsymbol{\gamma}$ é um vetor de coeficientes, $\mathbf{X}$ é uma matriz de variáveis explicativas, tais como escolaridade, experiência, cor, sexo, etc, e $\boldsymbol{\epsilon}$ é um vetor de erros.

Usualmente, os erros seguem uma distribuição normal e o rendimento-hora é computado da seguinte forma:

$$\text{rendimento-hora} = \text{rendimento mensal} / (\text{horas trabalhadas} \times 4.33).$$

Conforme descrito nos capítulos anteriores, a partir do modelo base de equação de rendimento minceriana (4.1), identifica-se que há a possibilidade de estudar rendimentos e horas trabalhadas separadamente para captar efeitos de remuneração (rendimento) e oferta de trabalho (horas trabalhadas), ou seja, tais variáveis podem captar comportamentos distintos. Adicionalmente, como tais variáveis tendem a ter correlação positiva, espera-se que o uso da regressão bivariada seja mais adequado nesse contexto.

Na estrutura de regressão bivariada, este modelo pode ser estimado na forma de vetor de variáveis dependentes $\mathbf{Y}_i = (Y_{1i}, Y_{2i})^\top$, onde Y_{1i} é o rendimento mensal no trabalho principal e Y_{2i} representa as horas semanais dedicadas ao trabalho principal para cada indivíduo i . Esse vetor pode ser modelado por meio de um conjunto $\mathbf{x}$ de variáveis explicativas e utilizando uma das três distribuições bivariadas explicadas no capítulo anterior, usando, respectivamente os conjuntos de equações de modo que:

- Normal

$$\begin{aligned} E[\log(Y_{1i})|\mathbf{X}_i = \mathbf{x}_i] &= \mu_{1i} = \mathbf{x}_i^\top \boldsymbol{\beta}_1, \quad i = 1, \dots, n, \\ E[\log(Y_{2i})|\mathbf{X}_i = \mathbf{x}_i] &= \mu_{2i} = \mathbf{x}_i^\top \boldsymbol{\beta}_2, \quad i = 1, \dots, n, \end{aligned} \quad (4.2)$$

- t

$$\begin{aligned} E[\log(Y_{1i})|\mathbf{X}_i = \mathbf{x}_i] &= \mu_{1i} = \mathbf{x}_i^\top \boldsymbol{\beta}_1, \quad i = 1, \dots, n, \\ E[\log(Y_{2i})|\mathbf{X}_i = \mathbf{x}_i] &= \mu_{2i} = \mathbf{x}_i^\top \boldsymbol{\beta}_2, \quad i = 1, \dots, n, \end{aligned} \quad (4.3)$$

- BS

$$\begin{aligned} E[Y_{1i}|\mathbf{X}_i = \mathbf{x}_i] &= \mu_{1i} = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_1), \quad i = 1, \dots, n, \\ E[Y_{2i}|\mathbf{X}_i = \mathbf{x}_i] &= \mu_{2i} = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_2), \quad i = 1, \dots, n. \end{aligned} \quad (4.4)$$

Nota-se que no caso da normal e t o logaritmo das variáveis dependentes é tomado, algo não necessário no caso BS. O uso do log tende a fornecer simetria, característica importante para o uso adequado do modelo normal e t .

A seguir, é feita a divisão das variáveis independentes utilizadas no modelo e a partir daí entender quais são os fatores que influenciam na composição e determinação da renda e das horas ofertadas. Para tal, é feita a separação das variáveis que afetam tanto o rendimento quanto as horas trabalhadas simultaneamente, assim como os fatores que afetam a um elemento e não ao outro. Tal escolha é feita através da literatura pertinente, Borjas (2012).

As variáveis independentes comuns que afetam tanto o rendimento quanto as horas trabalhadas são:

- Sexo - é uma dummy e apresenta valor 1-Homem ou 0-Mulher;
- Raça - é também uma dummy e apresenta valor 1-Brancos ou 0-Não brancos;
- Estado Civil - é uma dummy e possui valor 1-Casado ou 0-Não-Casado;
- Idade e Idade² - é incluída na equação a variável “idade” e seu respectivo valor ao quadrado, que é usada como proxy da experiência;
- Educação - classificada por anos formais de estudo, que variar entre 0 a 16 anos;
- Categoria (Alta, Média-Alta, Média, Média-Baixa, Baixa) - conjunto de binárias para captar a categoria de ocupação, segmentadas segundo critérios sócioeconômicos (tendo a categoria Baixa como referência)¹;
- Contrato (Com-Carteira, Sem-Carteira, Conta-Própria, Funcionário-Público) - as variáveis “sem carteira”, “conta própria” e “funcionário público” são dummies que buscam captar a posição na ocupação do indivíduo, tendo como referência os indivíduos com carteira;
- UFs (Pará, Ceará, Pernambuco, Bahia, Minas Gerais, Rio de Janeiro, Paraná, Rio Grande do Sul, Distrito Federal, São Paulo) - conjunto de binárias que captam as regiões metropolitanas, tomando São Paulo como referência;
- Ano (2013, 2014 e 2015) - são as dummies dos anos analisados, 2014 e 2015, tendo como referência o ano de 2013;

¹ Classificação ocupacional tem como base Jannuzzi (2001).

- Setor (Agrícola; Atividades industriais; Construção; Comércio, Alimentação e outros; Educação, saúde e serviços sociais)- são dummies que buscam captar o grupamento de atividade do indivíduo, tendo como referência os indivíduos que trabalham no setor público.

Em particular, as variáveis independentes que afetam apenas o rendimento são:

- Sindicato - é dummy e possui valor 1 para indivíduos que eram associados a algum sindicato no mês de referência e 0 para aqueles que não eram associados;
- Previdência - é dummy e possui valor 1 para indivíduos que eram contribuintes de algum tipo de previdência no mês de referência e 0 para aqueles que não eram contribuintes;
- Experiência - número de anos no trabalho principal, e pode variar entre 0 a 56 anos.

Por fim, as variáveis independentes que afetam apenas as horas trabalhadas são:

- Chefe - é uma dummy e está inserida com o intuito de captar a condição do indivíduo na unidade domiciliar, apresentando valor 1 para a pessoa de referência do domicílio (chefe) e valor 0 caso contrário (não-chefe);
- Menores - dummy usada para captar se existe crianças menores de 10 anos no domicílio;
- Inatividade - é uma dummy e possui valor 1 para domicílios que haviam desempregados no mês de referência e 0 para aqueles domicílios que todos estavam empregados.

Na seção seguinte é apresentado uma descrição sobre a interpretação dos coeficientes das regressões bivariadas consideradas nessa seção.

4.2.1 Interpretação dos coeficientes da regressão

Como mencionado anteriormente, no caso normal e t toma-se o logaritmo da variável dependente, e no caso BS tal variável permanece na sua escala original. No

entanto, é necessário entender como são interpretados os coeficientes de cada regressão. De fato, as interpretações são equivalentes para todos os modelos.

Seja $\log(Y_1)$ o log da variável rendimento (caso normal) na Equação (4.2). A interpretação do coeficiente da regressão para o l -ésimo regressor β_l é a taxa de mudança no $\log(Y_1)$ conforme X_l varia. Entretanto, tal interpretação não é muito útil uma vez que $\log(Y_1)$ muda de forma não linear com Y_1 . Uma interpretação alternativa é usada.

Considere o coeficiente β_{17} da primeira parte do modelo de regressão normal bivariada da Equação (4.2). Pode-se escrever

$$E[\log(Y_1)|X_7 = x_7, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}] = \mu_1 = \mathbf{X}^\top \boldsymbol{\beta}_1 = \beta_{10} + \beta_{11}x_1 + \mathbf{x}_{(7)}^\top \boldsymbol{\beta}_{1(7)},$$

em que $\mathbf{X}_{(7)}$, $\mathbf{x}_{(7)}$, $\boldsymbol{\beta}_{1(7)}$ são, respectivamente, os regressores excluindo $\mathbf{X}_7$, os valores observados dos regressores excluindo x_7 , e os coeficientes da regressão excluindo β_{17} . É possível fazer a aproximação do valor esperado do $\log(Y_1)$ pelo logaritmo do valor esperado, ou seja, $E[\log(Y_1)|X_7 = x_7, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}] \approx \log(E[Y_1|X_7 = x_7, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}])$, o qual implica

$$\begin{aligned} E[Y_1|X_7 = x_7, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}] &\approx \exp(E[\log(Y_1)|X_7 = x_7, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}]) \\ &= \exp(\beta_{10} + \beta_{17}x_7 + \mathbf{x}_{(7)}^\top \boldsymbol{\beta}_{1(7)}) \\ &= \exp(\beta_{17}x_7) \exp(\beta_{10} + \mathbf{x}_{(7)}^\top \boldsymbol{\beta}_{1(7)}). \end{aligned} \quad (4.5)$$

Note que se x_7 aumenta em 1 mantendo $\mathbf{X}_{(7)}$ constante, tem-se a multiplicação da média de Y_1 por aproximadamente $\exp(\beta_{17})$, ou seja,

$$\begin{aligned} E[Y_1|X_7 = x_7 + 1, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}] &\approx \exp(E[\log(Y_1)|X_7 = x_7 + 1, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}]) \\ &= \exp(\beta_{10} + \beta_{17}(x_7 + 1) + \mathbf{x}_{(7)}^\top \boldsymbol{\beta}_{1(7)}) \\ &= \exp(\beta_{17}(x_7 + 1)) \exp(\beta_{10} + \mathbf{x}_{(7)}^\top \boldsymbol{\beta}_{1(7)}) \\ &= \exp(\beta_{17}) \exp(\beta_{10} + \beta_{17}x_7 + \mathbf{x}_{(7)}^\top \boldsymbol{\beta}_{1(7)}) \\ &= \exp(\beta_{17}) E[Y_1|X_7 = x_7, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}] \\ &= \exp(\beta_{17}) \mu_1. \end{aligned} \quad (4.6)$$

As aproximações mostradas nas Equações (4.5) e (4.6) são importantes porque elas demonstram que a interpretação das estimativas dos parâmetros dos modelos de

regressão normal e t bivariados (Equações (4.2) e (4.3)) é a mesma do modelo BS (Equação (4.4)). Para mais detalhes ver Weisberg (2013).

Observa-se que a Equação (4.6) pode ser expressa em termos de percentagem. Em linhas gerais, o aumento (ou diminuição se o valor de β_{17} é negativo) percentual aproximado na variável dependente quando x_7 é aumentado em 1, é dado por

$$\left(\frac{E[Y_1|X_7 = x_7 + 1, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}] - E[Y_1|X_7 = x_7, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}]}{E[Y_1|X_7 = x_7, \mathbf{X}_{(7)} = \mathbf{x}_{(7)}]} \right) \times 100 = (\exp(\beta_{17}) - 1) \times 100. \quad (4.7)$$

Como exemplo, considerando X_7 educação (quantidade de anos estudados de 0 a 16) e supondo que $\beta_{17} = 0,2$, então $(\exp(\beta_{17}) - 1) \times 100 = 22,14\%$, ou um aumento de 22% no valor esperado do rendimento devido ao aumento de 1 em x_7 . Se $\beta_{17} = -0,2$, tem-se $(\exp(\beta_{17}) - 1) \times 100 = -18,13\%$, ou seja, o aumento em x_7 de 1, provoca uma diminuição de 18% no valor esperado do rendimento.

Para os casos em que o logaritmo natural é usado (como é o caso da aplicação dessa dissertação) e o valor de β_l é próximo de zero, digamos $-0,05 \leq \beta_l \leq 0,05$, pode-se aplicar a aproximação $(\exp(\beta_l) - 1) \approx \beta_l$. Em outras palavras, β_l é o aumento (ou diminuição se β_l é negativo) no valor esperado da variável dependente quando X_l aumenta em 1 e os outros regressores permanecem fixos. Assim, considerando X_7 educação e supondo $\beta_{17} = 0,0128$, tem-se um aumento de aproximadamente 1,3% no valor esperado do rendimento.

No caso de variáveis dummy, há uma interpretação em termos comparativos do valor esperado da variável dependente. Como exemplo, considere X_1 sexo (Homem=1 ou Mulher=0), então pode-se escrever

$$\begin{aligned} E[Y_1|X_1 = x_1, \mathbf{X}_{(1)} = \mathbf{x}_{(1)}] &\approx \exp(E[\log(Y_1)|X_1 = x_1, \mathbf{X}_{(1)} = \mathbf{x}_{(1)}]) \\ &= \exp(\beta_{10} + \beta_{11}x_1 + \mathbf{x}_{(1)}^\top \boldsymbol{\beta}_{1(1)}) \\ &= \exp(\beta_{10} + \beta_{11}x_1) \exp(\mathbf{x}_{(1)}^\top \boldsymbol{\beta}_{1(1)}), \end{aligned}$$

o que implica em

$$E[Y_1|X_1 = x_1, \mathbf{X}_{(1)} = \mathbf{x}_{(1)}] \approx \begin{cases} \exp(\beta_{10} + \beta_{11}) \exp(\mathbf{x}_{(1)}^\top \boldsymbol{\beta}_{1(1)}), & \text{if } x_1 = 1, \\ \exp(\beta_{10}) \exp(\mathbf{x}_{(1)}^\top \boldsymbol{\beta}_{1(1)}), & \text{if } x_1 = 0. \end{cases}$$

Logo, observa-se que o aumento (ou diminuição se o valor β_{11} é negativo) percentual aproximado na variável dependente quando x_1 muda de 0 para 1 é dado por

$$\left(\frac{E[Y_1|X_1 = 1, \mathbf{X}_{(1)} = \mathbf{x}_{(1)}] - E[Y_1|X_1 = 0, \mathbf{X}_{(1)} = \mathbf{x}_{(1)}]}{E[Y_1|X_1 = 0, \mathbf{X}_{(1)} = \mathbf{x}_{(1)}]} \right) \times 100 = (\exp(\beta_{11}) - 1) \times 100.$$

Similarmente, quando se tem o logaritmo natural é usado e o valor de β_l é próximo de zero ($-0,05 \leq \beta_l \leq 0,05$), aplicar-se a aproximação $(\exp(\beta_l) - 1) \approx \beta_l$. Dessa maneira, a suposição de que $\beta_{11} = 0.01$, sugere um aumento aproximado de 1% no valor esperado do salário dos homens em relação ao das mulheres.

O próximo capítulo apresentará uma análise exploratória dos dados utilizados, os resultados das estimações para cada modelo de regressão bivariada, e as comparações entre modelos.

5 Resultados dos modelos

Este capítulo visa apresentar e analisar os dados referentes ao estudo da estimação da equação de rendimentos minceriana em sua nova abordagem bivariada e também fazer uma comparação com a abordagem univariada utilizada até os dias atuais. A primeira seção apresenta as estatísticas descritivas sobre os dados analisados, já a segunda seção mostra a comparação dos modelos utilizando os critérios de informação e os gráficos de probabilidade-probabilidade (PP). Essa seção apresenta ainda os resultados das estimativas e suas interpretações.

5.1 Análise exploratória dos dados

Nessa seção é feita uma análise exploratória das variáveis dependentes (rendimento e horas trabalhadas). A Tabela 1 fornece algumas medidas descritivas para os dados do rendimento e horas trabalhadas em nível e na escala logarítmica, que incluem: tamanho amostral, mínimo, máximo, média, mediana, desvio-padrão (DP), coeficiente de variação (CV), coeficiente de assimetria (CA) e coeficiente de curtose (CC). Estas medidas indicam que o rendimento em nível apresenta assimetria elevada e curtose bastante acentuada, sugerindo que uma distribuição com assimetria e caudas pesadas seja mais adequada. Por outro lado, as horas trabalhadas em nível apresentam baixa assimetria e curtose moderada. A aplicação do logaritmo tende a trazer simetria, em especial para o rendimento. A Figura 1 mostra os histogramas do rendimento e horas trabalhadas em nível e na escala logarítmica.

Tabela 1 – Estatísticas descritivas para rendimento e horas trabalhadas (nível e escala logarítmica).

Data	<i>n</i>	Min.	Max.	Mediana	Média	DP	CV	CA	CC
Rendimento	167271	1.222	146677,6	1344.544	2278.51	3087.265	135.495%	7.056	116.514
Horas trab.	167271	1	98	40	39.909	11.554	28.951%	-0.550	3.445
log do Rendimento	167271	0.201	11.896	7.204	7.204	0,801	10.895%	0.409	2.008
log das horas trab.	167271	0.000	4.585	3.689	3.605	0.526	14.589%	-4.132	21.874

Fonte: elaboração própria a partir dos dados da PNAD.

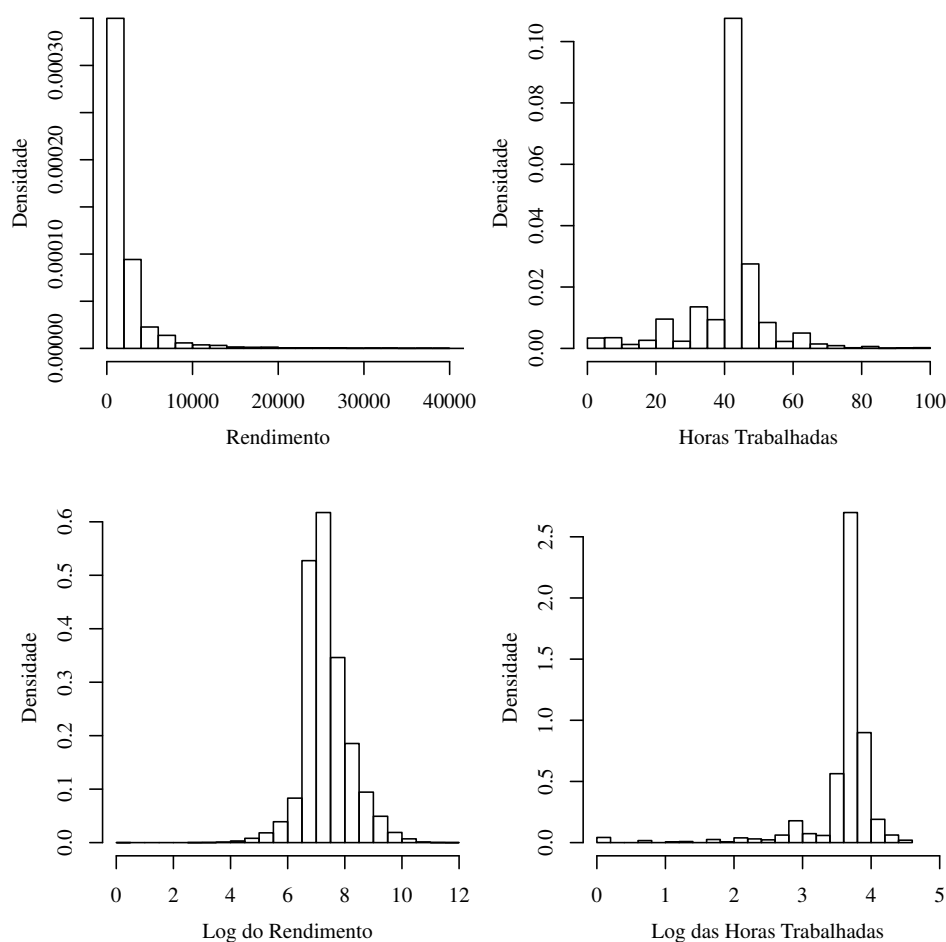


Figura 1 – Histograma para rendimento e horas trabalhadas (nível e escala logarítmica).

5.2 Resultados econométricos

Os resultados dessa seção são obtidos utilizando o software estatístico R. Parte dos códigos estão disponíveis no Apêndice B.

5.2.1 Investigação do melhor ajuste

Nessa seção, são apresentados os resultados das estimativas dos parâmetros dos modelos usando o método da máxima verossimilhança. Inicialmente são considerados os modelos de regressão univariada normal, t e BS para o rendimento e horas trabalhadas, e

depois suas versões bivariadas.

A Tabela 2 apresenta os valores dos critérios de informação Akaike (AIC) e Bayesiano (BIC), os quais são dados por

$$\text{AIC} = -2\ell + 2k \quad \text{e} \quad \text{BIC} = -2\ell + k \log(n),$$

em que ℓ é o valor da função de log-verossimilhança, k significa o número de parâmetros e n o número de observações. Os critérios de informação AIC e BIC são utilizados para seleção de modelos; quanto menor seus valores, melhor o modelo. É possível observar da Tabela 2, que os modelos tanto univariados quando bivariados baseados na t apresentam os melhores ajustes, pois apresentam os menores valores de AIC e BIC. Dessa forma dentre os três modelos, indica-se o uso dos modelos t univariado e bivariado.

Tabela 2 – Resultados das critérios de informação AIC e BIC para os modelos indicados.

		Modelos Univariados		
		Normal	t	BS
AIC	rendimento	272071.9	255565	2753495
	horas trab.	249342	60438.3	1539515
BIC	rendimento	272422.9	255936	2753846
	horas trab.	249692.9	60809.31	1539866
		Modelos Bivariados		
		Normal	t	BS
AIC	rendimento e horas trab.	515256.9	451026	4281716
BIC	rendimento e horas trab.	515968.8	451737.9	4282428

Fonte: elaboração própria a partir dos dados da PNAD.

Uma vez verificado os melhores modelos univariado e bivariado através dos valores AIC e BIC, procede-se a outra verificação do ajuste através do uso da distância de Mahalanobis - Marchant, Leiva e Cysneiros (2016). No caso da distribuição bivariada t , essa distância é dada por

$$D = \frac{1}{2}(\mathbf{U} - \boldsymbol{\mu})^\top \boldsymbol{\psi}^{-1}(\mathbf{U} - \boldsymbol{\mu}) \sim F_{2,\nu}, \quad (5.1)$$

em que $\mathbf{U} \sim t\text{Biv}(\mu_1, \mu_2, \sigma_1, \sigma_2, \nu, \rho)$ conforme Equação (3.16), e $\boldsymbol{\psi}$ é a matriz de covariância. De acordo com (5.1), a distância de Mahalanobis no caso da distribuição bivariada t segue uma distribuição $F_{2,\nu}$, ou seja, F com 2 e ν graus de liberdade;

no caso univariado tem-se uma $F_{1,\nu}$. Para a obtenção dos valores da distância de Mahalanobis, substituí-se os parâmetros pelos estimadores de máxima verossimilhança, o que resulta assintoticamente na mesma distribuição em (5.1) - Vilca, Balakrishnan e Zeller (2014). Pode-se então aplicar a aproximação de Wilson-Hilferty na distância Mahalanobis para obter uma distribuição aproximadamente normal padrão em (5.1). Desse modo, a qualidade do ajuste dos modelos de regressão univariados e bivariados t podem ser avaliados pelas distâncias transformadas com a aproximação de Wilson-Hilferty - Ibacache-Pulgar, Paula e Galea (2014). Nesses casos, as distâncias em (5.1) são adaptadas para acomodar a estrutura regressiva e a condição univariada ou bivariada.

A Figura 2 mostra os gráficos de probabilidade-probabilidade (PP) da distância Mahalanobis transformada para os modelos de regressão univariada t do rendimento e horas trabalhadas. O gráfico PP é comumente usado para avaliar o quão perto dois conjuntos de dados concordam, o qual é feito através do gráfico das duas correspondentes FDAs. Quanto mais próximo os pontos da linha de 45° de (0,0) a (1,1), melhor o ajuste. No caso da Figura 2, tem-se o gráfico da FDA teórica da normal padrão versus a FDA empírica da distância Mahalanobis transformada. A partir desse figura, observa-se o bom ajuste dos modelos univariados.

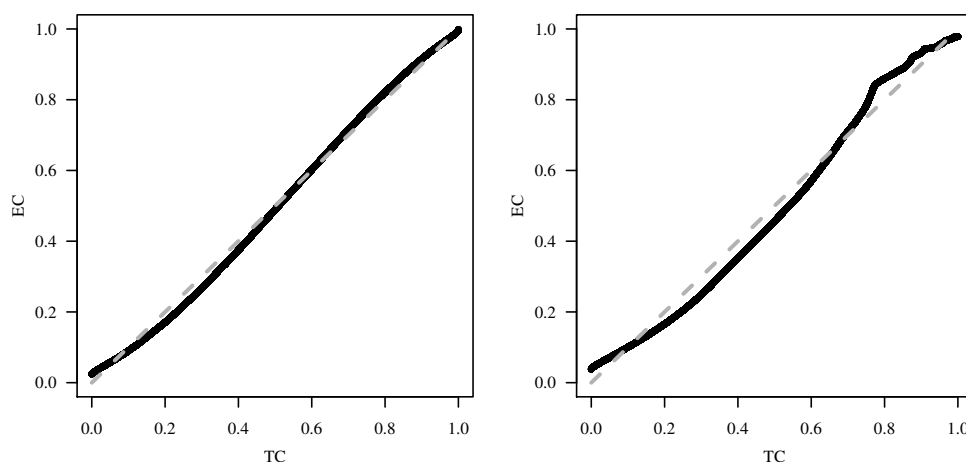


Figura 2 – Gráficos PP das distâncias transformadas para os modelos de regressão: univariada t do rendimento (esquerdo) e univariada t das horas trabalhadas (direito) - EC: probabilidade empírica; TC: probabilidade teórica.

A Figura 3 apresenta o gráfico PP da distância Mahalanobis transformada para o

modelo de regressão bivariada t do rendimento e horas trabalhadas. Nota-se que para o caso bivariado tem-se um bom ajuste. Os resultados sugerem assim que tanto no caso univariado quanto bivariado, os modelos t fornecem bons ajustes, e, portanto, podem ser utilizados.

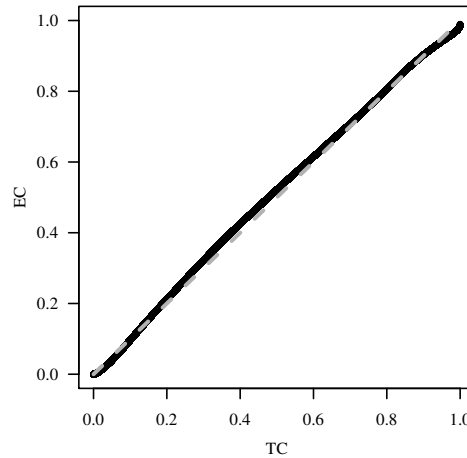


Figura 3 – Gráfico PP da distância transformada para o modelo de regressão bivariada t do rendimento e horas trabalhadas - EC: probabilidade empírica; TC: probabilidade teórica.

5.2.2 Resultados das estimativas

As Tabelas 3–5 apresentam os resultados das estimativas de máxima verossimilhança para os modelos de regressão t univariados do rendimento e horas trabalhadas, e o modelo de regressão bivariada t do rendimento e horas trabalhadas, com seus respectivos erros padrões, p valores e estatísticas Wald (W). Os modelos baseados na t tiveram os melhores ajustes e apenas são interpretados seus respectivos resultados. O Apêndice A apresenta os resultados para os casos normal e BS. Usa-se o teste de Wald para testar hipóteses de

$$H_0: \theta = \theta_0 \quad \text{versus} \quad H_1: \theta \neq \theta_0.$$

A estatística de Wald é definida por

$$W = \frac{[\hat{\theta} - \theta_0]}{\text{Erro Padrão}(\hat{\theta})},$$

o qual é aproximadamente distribuída como normal padrão $N(0,1)$ sob H_0 , em que $\hat{\theta}$ e θ_0 são o estimador e o seu respectivo valor proposto sob H_0 , respectivamente. No caso dessa dissertação, o interesse é $\theta_0 = 0$, ou seja,

$$H_0: \theta = 0 \quad \text{versus} \quad H_1: \theta \neq 0,$$

de maneira que a um nível de significância $\alpha = 0.05(5\%)$, tem-se que:

$$\begin{aligned} \text{se } p\text{-valor} > 0.05 &\rightarrow \text{não se rejeita } H_0, \\ \text{se } p\text{-valor} \leq 0.05 &\rightarrow \text{rejeita } H_0. \end{aligned}$$

No que tange a interpretação das estimativas dos coeficientes, apresentados na Seção 4.2.1, temos basicamente dois casos:

- Quando a variável dependente x é quantitativa (por exemplo, educação) e o valor do coeficiente: (a) está fora do intervalo $-0,05 \leq \hat{\beta} \leq 0,05$, há um aumento (ou diminuição se a estimativa é negativa) de $(\exp(\hat{\beta}) - 1) \times 100\%$ no valor esperado (média) da variável dependente devido a um aumento de 1 em x ; (b) está dentro do intervalo $-0,05 \leq \hat{\beta} \leq 0,05$, há um aumento (ou diminuição se a estimativa é negativa) $(\exp(\hat{\beta}) - 1) \approx \hat{\beta} \times 100\%$ no valor esperado (média) da variável dependente quando x aumenta em 1;
- Quando a variável dependente x é dicotômica (por exemplo, sexo) e o valor do coeficiente: (a) está fora do intervalo $-0,05 \leq \hat{\beta} \leq 0,05$, há um aumento (ou diminuição se a estimativa é negativa) de $(\exp(\hat{\beta}) - 1) \times 100\%$ no valor esperado (média) da variável dependente quando x muda de 0 (mulher) para 1 (homem); (b) está dentro do intervalo $-0,05 \leq \hat{\beta} \leq 0,05$, tem-se um aumento (ou diminuição se a estimativa é negativa) $(\exp(\hat{\beta}) - 1) \approx \hat{\beta} \times 100\%$ no valor esperado (média) da variável dependente quando x muda de 0 (mulher) para 1 (homem).

5.2.2.1 Estimativas univariadas t

Uma vez que os resultados das regressões univariadas do rendimento e horas trabalhadas com distribuição t não são o foco dessa dissertação, será comentado brevemente alguns dos resultados das estimações referentes a esses modelos, Tabelas 3 e 4.

Considerando os resultados para o modelo de regressão univariada do rendimento com a distribuição t , Tabela 3, a variável “Sexo” indica que homens possuem rendimento médio 29,69% maior que mulheres. Por outro lado, indivíduos possuem incremento no rendimento médio de 3,95% para cada ano acrescido na “Educação”. Importante notar que indivíduos que contribuem para a previdência têm rendimento médio 40.49% maior que indivíduos que não contribuem. As variáveis “Estado Civil” e “2014” não são estatisticamente significativas a um nível de 5%.

Tabela 3 – Estimativas para o modelo de regressão univariada do rendimento com a distribuição t ($\nu = 4$).

Variável dependente	Variável independente	Estimativa	Erro Padrão	Valor W	P-valor
Rendimento	(Intercept)	5.8886	0.0160	367.36	<0.0001
	Sexo	0.2600	0.0026	100.64	<0.0001
	Idade	0.0443	0.0006	68.41	<0.0001
	Idade ²	-0.0005	0.0001	-59.65	<0.0001
	Raça	0.0802	0.0026	30.50	<0.0001
	Estado Civil	0.0017	0.0059	0.29	0.7682
	Educação	0.0395	0.0004	99.77	<0.0001
	Pará	-0.3136	0.0054	-58.40	<0.0001
	Ceará	-0.3298	0.0051	-65.18	<0.0001
	Pernambuco	-0.3285	0.0049	-67.16	<0.0001
	Bahia	-0.2902	0.0049	-59.06	<0.0001
	Minas Gerais	-0.0879	0.0047	-18.76	<0.0001
	Rio de Janeiro	-0.0858	0.0043	-19.76	<0.0001
	Paraná	0.0240	0.0056	4.28	<0.0001
	Rio Grande do Sul	-0.0757	0.0044	-17.21	<0.0001
	Distrito Federal	0.0779	0.0053	14.77	<0.0001
	2014	0.0046	0.0028	1.62	0.1049
	2015	-0.0484	0.0029	-16.93	<0.0001
	Alta	0.8813	0.0068	128.76	<0.0001
	Média-alta	0.3292	0.0062	53.09	<0.0001
	Média	0.1584	0.0059	27.02	<0.0001
	Média-baixa	0.0262	0.0057	4.58	<0.0001
	Agrícola	-0.3960	0.0141	-28.05	<0.0001
	Atividades industriais	-0.2074	0.0071	-29.21	<0.0001
	Construção	-0.1047	0.0076	-13.69	<0.0001
	Comércio, alimentação e outros	-0.2311	0.0067	-34.51	<0.0001
	Educação, saúde e serviços sociais	-0.2638	0.0063	-41.68	<0.0001
	Outros serviços	-0.2340	0.0067	-35.19	<0.0001
	Sem-Carteira	-0.1621	0.0035	-46.93	<0.0001
	Conta-Própria	-0.0866	0.0033	-26.35	<0.0001
	Funcionário-Público	0.2597	0.0059	44.28	<0.0001
	Sindicato	0.1148	0.0033	34.95	<0.0001
	Previdencia	0.3407	0.0057	59.57	<0.0001
	Experiência	0.0123	0.0002	70.85	<0.0001

Fonte: elaboração própria a partir dos dados da PNAD.

A Tabela 4 apresenta os resultados para o modelo de regressão univariada das horas trabalhadas com a distribuição t . Neste caso, é feito o estudo do impacto das

variáveis independentes sobre o valor médio das horas trabalhadas, ou seja, o impacto sobre a oferta de trabalho. Percebe-se que a variável “Sexo” indica que homens possuem oferta média de horas trabalhadas 3,39% maior que das mulheres. A variável “Educação” tem um impacto negativo e pequeno sobre o as horas trabalhadas média. As variáveis “Raça”, “2014”, “Outros serviços”, “Menores” e “Desempregado” não são estatisticamente significativas a um nível de 5%.

Tabela 4 – Estimativas para o modelo de regressão univariada das horas trabalhadas com a distribuição t ($\nu = 4$).

Variável dependente	Variável independente	Estimativa	Erro Padrão	Valor W	P-valor
Horas trabalhadas	(Intercept)	3.5850	0.0069	522.79	<0.0001
	Sexo	0.0339	0.0011	30.42	<0.0001
	Idade	0.0057	0.0003	20.20	<0.0001
	Idade ²	-0.0001	0.0001	-20.12	<0.0001
	Raça	0.0004	0.0011	0.36	0.7200
	Estado Civil	0.0076	0.0025	3.00	0.0027
	Educacao	-0.0015	0.0002	-8.71	<0.0001
	Pará	-0.0320	0.0023	-13.98	<0.0001
	Ceará	-0.0022	0.0022	-1.01	0.3110
	Pernambuco	-0.0247	0.0021	-11.83	<0.0001
	Bahia	-0.0242	0.0021	-11.54	<0.0001
	Minas Gerais	-0.0185	0.0020	-9.26	<0.0001
	Rio de Janeiro	-0.0204	0.0019	-10.99	<0.0001
	Paraná	-0.0226	0.0024	-9.41	<0.0001
	Rio Grande do Sul	-0.0046	0.0019	-2.45	0.0144
	Distrito Federal	-0.0263	0.0022	-11.68	<0.0001
	2014	-0.0085	0.0012	-7.08	<0.0001
	2015	-0.0162	0.0012	-13.23	<0.0001
	Alta	0.0368	0.0029	12.70	<0.0001
	Média-alta	0.0269	0.0026	10.18	<0.0001
	Média	0.0574	0.0025	23.01	<0.0001
	Média-baixa	0.0604	0.0024	24.79	<0.0001
	Agrícola	0.0913	0.0060	15.20	<0.0001
	Atividades industriais	0.0217	0.0030	7.16	<0.0001
	Construção	0.0236	0.0033	7.25	<0.0001
	Comércio, alimentação e outros	0.0564	0.0029	19.76	<0.0001
	Educação, saúde e serviços sociais	-0.0276	0.0027	-10.22	<0.0001
	Outros serviços	0.0040	0.0028	1.40	0.1613
	Sem-Carteira	-0.0766	0.0015	-52.47	<0.0001
	Conta-Própria	-0.0629	0.0014	-46.20	<0.0001
	Funcionário-Público	-0.0383	0.0025	-15.49	<0.0001
	Chefe	0.0091	0.0011	8.41	<0.0001
	Menores	0.0018	0.0011	1.61	0.1065
	Inatividade	0.0016	0.0016	1.03	0.3039

Fonte: elaboração própria a partir dos dados da PNAD.

5.2.2.2 Estimativas bivariadas t e comparação com equação minceriana tradicional

A Tabela 5 apresenta os resultados para o modelo de regressão bivariada do rendimento e horas trabalhadas com a distribuição t bivariada. E, para algumas variáveis, são similares aos do caso univariado para cada variável dependente. Nota-se que o coeficiente de correlação estimado entre rendimento e horas trabalhadas é de 0,1897. E, também, que o coeficiente de correlação é estatisticamente significativo a um nível de 5%, ou seja, o modelo bivariado é mais adequado do que estimar as equações de forma separada. Desse modo, considerar equações separadas pode levar a predições incorretas.

Considerando os resultados para o modelo de regressão bivariada do rendimento e as horas trabalhadas com a distribuição t , Tabela 5, a variável “Sexo” indica que homens possuem renda média 34,58% maior que mulheres, e que ofertam 8,62% de horas trabalhadas a mais em média do que as mulheres - o último resultado está em conformidade com o encontrado em Gonzaga, Leite e Machado (2002). Já a variável “Raça”, mostra que os brancos recebem em média 10,31% a mais do que os não-brancos, mas se tratando das horas trabalhadas os brancos ofertam apenas 0,02% a mais do que os não-brancos. Este resultado também confirma trabalhos anteriores, no qual é discutido a hipótese de discriminação no mercado de trabalho. Como mostrado em Cavalieri e Fernandes (1998), os autores confirmam essa hipótese utilizando dados da PNAD 1989, e constata que, em média, tanto a renda dos homens são superiores aos das mulheres como a renda dos indivíduos de cor branca são superiores aos de cor não-branca; mesmo após a realização de uma série de controles, como: idade, anos de estudo e região de residência.

A variável “Idade” indica que para cada ano adicional, há um aumento de 5,27% na renda média e que apenas 1,28% é acrescido na média das horas trabalhadas. No que concerne a variável “Educação”, quando há um aumento de um ano de estudo, há também um aumento na renda média de 4,68%, porém se tratando das horas trabalhadas, há um decréscimo de magnitude 0,09%, ou seja, quanto mais anos acrescidos de estudos, maior a renda média, porém menor a quantidade de horas ofertadas de trabalho. Há, aqui, um ponto fundamental na vantagem de se considerar a regressão bivariada, uma vez que o efeito dessa variável é distinto nas duas equações, efeito que não é evidenciado pelo modelo tradicional de regressão univariada que toma o rendimento-hora como variável

dependente. Especificamente, no que se refere ao resultado econômico da variável “Educação”, Lau et al. (1993) encontram também uma estimativa positiva do aumento do rendimento (*per capita*) em decorrência do aumento dos anos de estudo. O que reforça os pressupostos da teoria do capital humano, que discute o nível educacional dos indivíduos como fator explicativo para o aumento na renda. No tocante às horas trabalhadas, Gonzaga, Leite e Machado (2002) afirmam que, no Brasil, os mais escolarizados trabalham menos.

Sobre as regiões metropolitanas, é observado que o Distrito Federal, por exemplo, possui uma renda média 8,95% maior e oferta 2,70% a menos de horas trabalhadas, tomando São Paulo como referência. Ou seja, aqui, também, é possível identificar efeitos distintos (sinais diferentes), e que tal efeito não é passível de captação no modelo tradicional univariado. Para explicar tal resultado deve-se considerar as características não observáveis dos trabalhadores tais como habilidade, motivação, entre outras, e as diferenças específicas dos setores e das regiões. Como ressalta Fernandes (2002), segmentos distintos podem apresentar remunerações distintas.

Em relação aos tipos de contrato, que buscam captar a posição na ocupação do indivíduo, tomando os indivíduos na modalidade “Com Carteira” como referência, é observado que, aqueles “Sem Carteira” tem renda média 17,73% menor se comparado ao indivíduos que possuem carteira assinada e também ofertam menos horas trabalhadas, numa magnitude de 15,08% a menos. Se comparado a modalidade “Funcionário Público”, nota-se que, em média há um acréscimo de 26,90% na renda desses trabalhadores. Por outro lado, funcionários públicos ofertam menos horas trabalhadas, cerca de 4,10% a menos. Já aqueles que trabalham por “Conta Própria”, tem rendimento médio 12,82% inferior e ofertam 16,04% a menos de horas trabalhadas - no que se refere as horas dedicadas ao trabalho, todos esses resultados são condizentes com os encontrados em Gonzaga, Leite e Machado (2002). Porém, vale ressaltar que o ponto de interesse neste trabalho é o fato dos resultados mostrarem efeitos distintos, e aqui, novamente, há uma limitação do modelo univariado por não captar essa diferenciação de sinais, como o encontrado na variável “Funcionário Público”.

Ainda em relação aos resultados das estimativas da Tabela 5, percebe-se que, as variáveis “Estado Civil” e “2014” não são estatisticamente significativas a um nível de 5% para o rendimento, sendo que a variável “Estado Civil” também não é significativa

no que tange às horas trabalhadas.

Destacando agora as variáveis que afetam apenas o rendimento ou apenas as horas trabalhadas, a Tabela 5 indica que a variável "Previdência", que afeta apenas o rendimento, tem como resultado 42,70%, ou seja, indivíduos que contribuem para a previdência tem uma renda média 42,70% maior do que aqueles indivíduos que não contribuem. Já a variável "Chefe", que afeta apenas as horas trabalhadas, confirma que chefes de família ofertam 1,80% horas a mais de trabalho em média.

A título de comparação, é apresentando na Tabela 6 os resultados da regressão univariada tradicional do rendimento-hora. Ou seja, aqui os resultados consideram como variável dependente o rendimento-hora, o qual é a variável dependente tradicionalmente usada no modelo minceriano. A princípio, alguns resultados apresentam similaridade, em termos de magnitude, aos resultados do modelo de regressão bivariado. No entanto, como destacado anteriormente, o modelo univariado por definição não consegue separar efeitos distintos, cuja variáveis independentes têm sobre rendimento e oferta de trabalho. Pode-se destacar, em especial, três variáveis "Educação", "Distrito Federal" e "Funcionário Público". Na regressão bivariada, tais variáveis possuem sinais distintos para rendimento e horas trabalhadas, por exemplo, considerando a variável "Educação" quando há um aumento de um ano de estudo, há também um aumento na renda média de 4,68%, já sobre as horas trabalhadas há um decréscimo de magnitude 0,09%, isso mostra que quanto mais anos acrescidos de estudos, maior a renda média, porém menos horas de trabalho são ofertadas. Por outro lado, na regressão univariada apresentada na Tabela 6, só é observado o impacto positivo de um ano adicional no rendimento-hora médio, 4,79%. Desse modo, não é possível captar o efeito duplo (positivo e negativo) no modelo de regressão univariada. Além disso, o modelo de regressão bivariada capta a correlação entre as duas variáveis dependentes, o que no caso da presente dissertação implica em um modelo preditivo mais acurado, uma vez que a correlação no modelo bivariado é positiva e significativa. Portanto, há notórias vantagens do modelo bivariado sobre o univariado, sobretudo nos próprios resultados das estimativas, como mostrado nas análises dessa seção. Assim, o modelo bivariado mostra-se mais adequado para a modelagem da equação minceriana.

Tabela 5 – Estimativas para o modelo de regressão bivariada do rendimento e horas trabalhadas com a distribuição t ($\nu = 4$) bivariada.

Variável depen.	Variável indep.	Estimativa	Erro Padrão	Valor W	P-valor
Rendimento	σ_1	0.5456	0.0019	294.0464	<0.0001
	σ_2	0.5098	0.0018	292.0464	<0.0001
	(Intercept)	5.6253	0.0194	290.4145	<0.0001
	Sexo	0.297	0.0031	96.2507	<0.0001
	Idade	0.0514	8e-04	66.7815	<0.0001
	Idade ²	-6e-04	8e-06	-57.561	<0.0001
	Raça	0.0982	0.0031	31.8546	<0.0001
	Estado Civil	0.0065	0.0069	0.935	0.3498
	Educação	0.0468	5e-04	97.6625	<0.0001
	Pará	-0.3187	0.0063	-50.8093	<0.0001
	Ceará	-0.3502	0.0059	-59.2043	<0.0001
	Pernambuco	-0.3498	0.0057	-61.2782	<0.0001
	Bahia	-0.312	0.0058	-53.8771	<0.0001
	Minas Gerais	-0.0931	0.0055	-17.0357	<0.0001
	Rio de Janeiro	-0.0794	0.0052	-15.3409	<0.0001
	Paraná	0.0121	0.0065	1.85	0.0643
	Rio Grande do Sul	-0.0884	0.0051	-17.2281	<0.0001
	Distrito Federal	0.0858	0.0063	13.571	<0.0001
	2014	0.0024	0.0033	0.7242	0.469
	2015	-0.0589	0.0034	-17.5054	<0.0001
	Alta	0.8855	0.0084	105.5256	<0.0001
	Média-alta	0.3708	0.0073	50.5346	<0.0001
	Média	0.1862	0.0069	26.8458	<0.0001
	Média-baixa	0.0507	0.0068	7.4931	<0.0001
	Agrícola	-0.4309	0.018	-23.969	<0.0001
	Atividades industriais	-0.2036	0.0087	-23.3776	<0.0001
	Construção	-0.0781	0.0093	-8.3839	<0.0001
	Comércio, alimentação e outros	-0.2107	0.0083	-25.4561	<0.0001
	Educação, saúde e serviços sociais	-0.2543	0.0079	-32.1732	<0.0001
	Outros serviços	-0.214	0.0083	-25.8597	<0.0001
	Sem-Carteira	-0.1952	0.0041	-47.2089	<0.0001
	Conta-Própria	-0.1373	0.0042	-32.6763	<0.0001
	Funcionário-Público	0.2383	0.0072	33.2859	<0.0001
	Sindicato	0.122	0.0038	32.4575	<0.0001
	Previdência	0.3556	0.0072	49.3444	<0.0001
	Experiência	0.0127	2e-04	59.777	<0.0001
Horas trabalhadas	(Intercept)	3.2761	0.0166	196.993	<0.0001
	Sexo	0.0827	0.0027	30.554	<0.0001
	Idade	0.0128	7e-04	18.7405	<0.0001
	Idade ²	-1e-04	0	-15.1547	<0.0001
	Raça	2e-04	0.0027	0.0855	0.9318
	Estado Civil	-0.0092	0.0061	-1.5092	0.1312
	Educação	-9e-04	4e-04	-2.2331	0.0255
	Pará	0.0036	0.0055	0.6579	0.5106
	Ceará	0.0172	0.0052	3.3292	9e-04
	Pernambuco	-0.0149	0.005	-2.9866	0.0028
	Bahia	-0.044	0.0051	-8.6772	<0.0001
	Minas Gerais	0.012	0.0047	2.5458	0.0109
	Rio de Janeiro	-0.0874	0.0045	-19.5936	<0.0001
	Paraná	-0.013	0.0057	-2.2867	0.0222
	Rio Grande do Sul	0.0171	0.0044	3.8516	1e-04
	Distrito Federal	-0.027	0.0054	-5.0099	<0.0001
	2014	0.0119	0.0029	4.142	<0.0001
	2015	-0.0394	0.0029	-13.4526	<0.0001
	Alta	0.123	0.0073	16.8922	<0.0001
	Média-alta	0.0997	0.0066	15.067	<0.0001
	Média	0.1465	0.0063	23.2992	<0.0001
	Média-baixa	0.155	0.0061	25.272	<0.0001
	Agrícola	0.1566	0.0152	10.3171	<0.0001
	Atividades industriais	0.0204	0.0072	2.8169	0.0048
	Construção	0.0426	0.0078	5.4722	<0.0001
	Comércio, alimentação e outros	0.0651	0.0069	9.4892	<0.0001
	Educação, saúde e serviços sociais	-0.0658	0.0065	-10.0724	<0.0001
	Outros serviços	-0.0147	0.0069	-2.1406	0.0323
	Sem-Carteira	-0.1635	0.0036	-45.2849	<0.0001
	Conta-Própria	-0.1749	0.0034	-50.8516	<0.0001
	Funcionário-Público	-0.041	0.006	-6.856	<0.0001
	Chefe	0.018	0.0027	6.7189	<0.0001
	Menores	-0.0028	0.0026	-1.1009	0.2709
	Inatividade	-0.0055	0.0037	-1.4954	0.1348
	ρ	0.1897	0.0131	14.5346	<0.0001

Fonte: elaboração própria a partir dos dados da PNAD.

Tabela 6 – Estimativas para o modelo de regressão univariada do rendimento-hora tradicional.

Variável dependente	Variável independente	Estimativa	Erro Padrão	Valor W	P-valor
Rendimento-hora	(Intercept)	0.9428	0.0229	41.23	<0.0001
	Sexo	0.2027	0.0037	54.42	<0.0001
	Idade	0.0347	0.0009	36.99	<0.0001
	Idade ²	-0.0004	0.0001	-30.31	<0.0001
	Raça	0.1005	0.0037	26.91	<0.0001
	Estado Civil	0.0097	0.0084	1.16	0.2477
	Educação	0.0479	0.0006	85.13	<0.0001
	Pará	-0.3155	0.0076	-41.37	<0.0001
	Ceará	-0.3673	0.0072	-51.15	<0.0001
	Pernambuco	-0.3321	0.0069	-47.83	<0.0001
	Bahia	-0.2674	0.0070	-38.33	<0.0001
	Minas Gerais	-0.1046	0.0067	-15.73	<0.0001
	Rio de Janeiro	0.0060	0.0062	0.97	0.3332
	Paraná	0.0205	0.0080	2.57	0.0101
	Rio Grande do Sul	-0.1097	0.0062	-17.58	<0.0001
	Distrito Federal	0.1140	0.0075	15.23	<0.0001
	2014	-0.0087	0.0040	-2.15	0.0312
	2015	-0.0159	0.0041	-3.90	0.0001
	Alta	0.7637	0.0097	78.59	<0.0001
	Média-alta	0.2750	0.0088	31.24	<0.0001
	Média	0.0399	0.0083	4.79	<0.0001
	Média-baixa	-0.1060	0.0081	-13.04	<0.0001
	Agrícola	-0.5832	0.0200	-29.11	<0.0001
	Atividades industriais	-0.2242	0.0101	-22.24	<0.0001
	Construção	-0.1194	0.0109	-11.00	<0.0001
	Comércio, alimentação e outros	-0.2780	0.0095	-29.25	<0.0001
	Educação, saúde e serviços sociais	-0.1875	0.0090	-20.87	<0.0001
	Outros serviços	-0.2004	0.0094	-21.24	<0.0001
	Sem-Carteira	-0.0343	0.0049	-7.00	<0.0001
	Conta-Própria	0.0397	0.0047	8.51	<0.0001
	Funcionário-Público	0.2880	0.0083	34.59	<0.0001
	Sindicato	0.1199	0.0047	25.71	<0.0001
	Previdência	0.3318	0.0081	40.86	<0.0001
	Experiência	0.0104	0.0002	42.06	<0.0001
	Chefe	0.0644	0.0036	17.92	<0.0001
	Menores	0.0122	0.0036	3.35	0.0008
	Inatividade	-0.0618	0.0052	-11.93	<0.0001

Fonte: elaboração própria a partir dos dados da PNAD.

6 Considerações finais

Nessa dissertação, a equação minceriana de rendimentos foi estudada e uma nova abordagem para tal equação foi proposta. A princípio, foi feito um estudo sobre sua origem, com uma breve revisão da literatura mostrando os diversos trabalhos empíricos em economia que têm como base essa equação. Dado o contexto em que os trabalhos abordam a equação minceriana, ou seja, tomando o logaritmo do rendimento-hora como variável dependente em um modelo de regressão linear univariado, observou-se que é possível dar uma nova abordagem a essa equação. Em especial, verificou-se que a junção de dois componentes, rendimento e horas trabalhadas, em apenas uma variável dependente (rendimento-hora) impossibilita a captação de efeitos distintos das duas variáveis, ou seja, recompensa ou prêmio pelo esforço despendido pela mão de obra, e a disponibilidade de tempo que o trabalhador disponibiliza ao mercado (oferta de trabalho). Desse modo, foi proposto a modelagem de regressão bivariada, separando a variável dependente rendimento-hora em dois componentes (rendimento e horas trabalhadas), de maneira a captar os distintos efeitos e a modelar também a correlação entre essas variáveis.

Na modelagem bivariada, foram usados os modelos de regressão bivariadas normal, t e Birnbaum-Saunders para modelar conjuntamente o rendimento e horas trabalhadas. Os resultados obtidos mostraram que os modelos de regressão t bivariado apresentaram os melhores ajustes para os dados considerados. Adicionalmente, observou-se uma correlação positiva e significativa entre rendimento e horas trabalhadas, ou seja, há uma justificativa para o uso da regressão bivariada em detrimento da estimação das regressões separadas. Um ponto importante, é que o uso de regressões separadas na presença de correlação leva a previsões incorretas, ou seja, a regressão bivariada é a mais indicada nesse contexto.

Por fim, a hipótese inicial deste trabalho, de que ambas as variáveis captam efeitos distintos foi confirmada através dos dados, pois, ao se analisar, por exemplo, a variável “Educação”, o estudo revelou que, quando há um aumento de um ano de estudo, há também um aumento na renda média de 4,68%, e há um decréscimo de magnitude

0,09% para as horas trabalhadas. Isso identificou que quanto mais anos acrescidos de estudos, maior a renda média, porém menos horas são ofertadas de trabalho, um ponto fundamental sobre a vantagem de se considerar a regressão bivariada, pois há um efeito distinto (sinais diferentes) da “Educação” na renda e na oferta de horas trabalhadas. Nota-se que tal efeito não pode ser captado pelo modelo tradicional de regressão univariada que toma o rendimento-hora como variável dependente.

Desse modo, espera-se que os resultados obtidos nessa dissertação sirvam como abordagem alternativa para a modelagem da equação minceriana de rendimentos. Como trabalhos futuros, pode-se implementar uma correção de viés de seleção (HECKMAN, 1979), uma vez os dados da PNAD se referem as pessoas que tinham trabalho nos anos da pesquisa. Entretanto, o rendimento do indivíduo está associado à decisão do mesmo trabalhar ou não, o que em última instância depende do custo de oportunidade - é vantajoso trabalhar se o rendimento recebido (ou rendimento potencial) é maior que o custo de oportunidade (rendimento reserva). Em adição, pode-se usar outras distribuições de probabilidade bivariadas para a modelagem do rendimento e horas trabalhadas, tal como a Pareto e suas extensões, comumente utilizadas para a modelagem da renda. Finalmente, um modelo de regressão logística bivariada poderia ser utilizado para estimar a influência das características individuais sobre as probabilidades de um determinado trabalhador pertencer a um particular grupo de renda e tipo de trabalho. Trabalhos sobre as mencionadas extensões estão em progresso.

A Resultados das estimativas

Esse apêndice apresenta os resultados das estimativas para os modelos de regressão univariado e bivariado normal e BS.

Tabela 7 – Estimativas para o modelo de regressão univariada do rendimento com a distribuição normal.

Variável dependente	Variável independente	Estimativa	Erro Padrão	Valor W	P-valor
Salário	(Intercept)	5.6253	0.0185	304.01	<0.0001
	Sexo	0.2970	0.0030	99.60	<0.0001
	Idade	0.0514	0.0007	68.63	<0.0001
	Idade ²	-0.0006	0.0001	-59.86	<0.0001
	Raça	0.0982	0.0030	32.36	<0.0001
	Estado Civil	0.0065	0.0068	0.95	0.3408
	Educação	0.0468	0.0005	102.35	<0.0001
	Pará	-0.3187	0.0062	-51.41	<0.0001
	Ceará	-0.3502	0.0058	-59.96	<0.0001
	Pernambuco	-0.3498	0.0056	-61.95	<0.0001
	Bahia	-0.3120	0.0057	-55.01	<0.0001
	Minas Gerais	-0.0931	0.0054	-17.22	<0.0001
	Rio de Janeiro	-0.0794	0.0050	-15.83	<0.0001
	Paraná	0.0121	0.0065	1.87	0.0619
	Rio Grande do Sul	-0.0884	0.0051	-17.41	<0.0001
	Distrito Federal	0.0858	0.0061	14.10	<0.0001
	2014	0.0024	0.0033	0.74	0.4614
	2015	-0.0589	0.0033	-17.85	<0.0001
	Alta	0.8855	0.0079	112.07	<0.0001
	Média-alta	0.3708	0.0072	51.80	<0.0001
	Média	0.1862	0.0068	27.52	<0.0001
	Média-baixa	0.0507	0.0066	7.67	<0.0001
	Agrícola	-0.4309	0.0163	-26.44	<0.0001
	Atividades industriais	-0.2036	0.0082	-24.84	<0.0001
	Construção	-0.0781	0.0088	-8.84	<0.0001
	Comércio, alimentação e outros	-0.2107	0.0077	-27.26	<0.0001
	Educação, saúde e serviços sociais	-0.2543	0.0073	-34.80	<0.0001
	Outros serviço	-0.2140	0.0077	-27.88	<0.0001
	Sem-Carteira	-0.1952	0.0040	-48.95	<0.0001
	Conta-Própria	-0.1373	0.0038	-36.21	<0.0001
	Funcionário-Público	0.2383	0.0068	35.19	<0.0001
	Sindicato	0.1220	0.0038	32.19	<0.0001
	Previdência	0.3556	0.0066	53.86	<0.0001
	Experiência	0.0127	0.0002	63.13	<0.0001

Fonte: elaboração própria a partir dos dados da PNAD.

Tabela 8 – Estimativas para o modelo de regressão univariada das horas trabalhadas com a distribuição normal.

Variável dependente	Variável independente	Estimativa	Erro Padrão	Valor W	P-valor
Horas trabalhadas	(Intercept)	3.2761	0.0173	188.95	<0.0001
	Sexo	0.0827	0.0028	29.32	<0.0001
	Idade	0.0128	0.0007	18.04	<0.0001
	Idade ²	-0.0002	0.0001	-18.06	<0.0001
	Raça	0.0002	0.0028	0.08	0.9354
	Estado Civil	-0.0092	0.0064	-1.44	0.1487
	Educação	-0.0009	0.0004	-2.19	0.0287
	Pará	0.0036	0.0058	0.62	0.5333
	Ceará	0.0172	0.0055	3.15	0.0016
	Pernambuco	-0.0149	0.0053	-2.83	0.0047
	Bahia	-0.0440	0.0053	-8.30	<0.0001
	Minas Gerais	0.0120	0.0051	2.38	0.0172
	Rio de Janeiro	-0.0874	0.0047	-18.66	<0.0001
	Paraná	-0.0130	0.0061	-2.15	0.0319
	Rio Grande do Sul	0.0171	0.0047	3.61	0.0003
	Distrito Federal	-0.0270	0.0057	-4.76	<0.0001
	2014	0.0119	0.0031	3.91	0.0001
	2015	-0.0394	0.0031	-12.76	<0.0001
	Alta	0.1230	0.0073	16.76	<0.0001
	Média-alta	0.0997	0.0067	14.93	<0.0001
	Média	0.1465	0.0063	23.21	<0.0001
	Média-baixa	0.1550	0.0062	25.15	<0.0001
	Agrícola	0.1566	0.0152	10.31	<0.0001
	Atividades industriais	0.0204	0.0077	2.67	0.0076
	Construção	0.0426	0.0082	5.17	<0.0001
	Comércio, alimentação e outros	0.0651	0.0072	9.03	<0.0001
	Educação, saúde e serviços sociais	-0.0658	0.0068	-9.65	<0.0001
	Outros serviços	-0.0147	0.0072	-2.05	0.0405
	Sem-Carteira	-0.1635	0.0037	-44.28	<0.0001
	Conta-Própria	-0.1749	0.0034	-50.85	<0.0001
	Funcionário-Público	-0.0410	0.0063	-6.55	<0.0001
	Chefe	0.0180	0.0027	6.60	<0.0001
	Menores	-0.0028	0.0028	-1.02	0.3068
	Inatividade	-0.0055	0.0039	-1.41	0.1591

Fonte: elaboração própria a partir dos dados da PNAD.

Tabela 9 – Estimativas para o modelo de regressão univariada do rendimento com a distribuição BS.

Variável dependente	Variável independente	Estimativa	Erro Padrão	Valor W	P-valor
Salário	(Intercept)	5.6587	0.0143	394.4082	<0.0001
	Sexo	0.3115	0.0031	99.2942	<0.0001
	Idade	0.0566	2e-04	242.0012	<0.0001
	Idade ²	-6e-04	0	-1428.5416	<0.0001
	Raça	0.0922	0.0032	28.7049	<0.0001
	Estado Civil	0.0249	0.0073	3.4223	6e-04
	Educação	0.052	5e-04	109.6533	<0.0001
	Pará	-0.3044	0.0066	-46.2611	<0.0001
	Ceará	-0.3506	0.0062	-56.9516	<0.0001
	Pernambuco	-0.341	0.006	-57.0609	<0.0001
	Bahia	-0.3322	0.0059	-55.9463	<0.0001
	Minas Gerais	-0.115	0.0057	-20.2292	<0.0001
	Rio de Janeiro	-0.0539	0.0053	-10.0889	<0.0001
	Paraná	0.0301	0.0069	4.3624	<0.0001
	Rio Grande do Sul	-0.0695	0.0054	-12.8868	<0.0001
	Distrito Federal	0.0947	0.0064	14.7601	<0.0001
	2014	-0.0017	0.0035	-0.4871	0.6262
	2015	-0.0513	0.0035	-14.6603	<0.0001
	Alta	0.8736	0.0083	105.7249	<0.0001
	Média-alta	0.3451	0.0075	45.771	<0.0001
	Média	0.1817	0.0071	25.4657	<0.0001
	Média-baixa	0.0643	0.007	9.1944	<0.0001
	Agrícola	-0.4566	0.0163	-27.9662	<0.0001
	Atividades industriais	-0.2165	0.0086	-25.1102	<0.0001
	Construção	-0.0644	0.0093	-6.9123	<0.0001
	Comércio, alimentação e outros	-0.2133	0.0081	-26.2556	<0.0001
	Educação, saúde e serviços sociais	-0.2521	0.0077	-32.9059	<0.0001
	Outros serviços	-0.2283	0.008	-28.4657	<0.0001
	Sem-Carteira	-0.2073	0.0042	-48.9024	<0.0001
	Conta-Própria	-0.2133	0.0038	-55.527	<0.0001
	Funcionário-Público	0.2204	0.0072	30.6494	<0.0001
	Sindicato	0.1273	0.0041	31.3235	<0.0001
	Previdência	0.3667	0.0068	54.0149	<0.0001
	Experiência	0.014	2e-04	68.2876	<0.0001
	δ	5.4646	0.0189	289.103	<0.0001

Fonte: elaboração própria a partir dos dados da PNAD.

Tabela 10 – Estimativas para o modelo de regressão univariada das horas trabalhadas com a distribuição BS.

Variável dependente	Variável independente	Estimativa	Erro Padrão	Valor W	P-valor
Horas trabalhadas	(Intercept)	3.2709	0.0162	202.4886	<0.0001
	Sexo	0.0788	0.0035	22.3184	<0.0001
	Idade	0.014	3e-04	52.5154	<0.0001
	Idade ²	-2e-04	0	-331.9222	<0.0001
	Raça	-0.0036	0.0036	-1.0212	0.3072
	Estado Civil	-0.0408	0.0078	-5.2369	<0.0001
	Educação	-3e-04	5e-04	-0.4756	0.6344
	Pará	0.1018	0.0074	13.7199	<0.0001
	Ceará	0.0929	0.0069	13.4143	<0.0001
	Pernambuco	0.0551	0.0067	8.2667	<0.0001
	Bahia	-0.0191	0.0066	-2.9064	0.0037
	Minas Gerais	0.1266	0.0065	19.4595	<0.0001
	Rio de Janeiro	-0.1803	0.0056	-32.4855	<0.0001
	Paraná	0.0532	0.0077	6.9075	<0.0001
	Rio Grande do Sul	0.0934	0.006	15.5685	<0.0001
	Distrito Federal	-0.0188	0.007	-2.662	0.0078
	2014	0.0586	0.0039	15.0987	<0.0001
	2015	-0.0767	0.0038	-20.076	<0.0001
	Alta	0.0969	0.0091	10.5978	<0.0001
	Média-alta	0.0713	0.0083	8.5893	<0.0001
	Média	0.1283	0.0078	16.362	<0.0001
	Média-baixa	0.1392	0.0077	18.1668	<0.0001
	Agrícola	0.1602	0.0193	8.3001	<0.0001
	Atividades industriais	0.0499	0.0098	5.0989	<0.0001
	Construção	0.0373	0.0105	3.558	4e-04
	Comércio, alimentação e outros	0.0735	0.0092	7.9926	<0.0001
	Educação, saúde e serviços sociais	-0.0455	0.0087	-5.2418	<0.0001
	Outros serviços	-0.0087	0.0091	-0.955	0.3396
	Sem-Carteira	-0.1302	0.0046	-28.1645	<0.0001
	Conta-Própria	-0.1455	0.0043	-34.0548	<0.0001
	Funcionário-Público	-0.0206	0.008	-2.5889	0.0096
	Chefe	0.0275	0.0034	8.06	<0.0001
	Menores	-1e-04	0.0034	-0.0202	0.9839
	Inatividade	-0.0106	0.0049	-2.1629	0.0305
	δ	4.3763	0.0151	288.977	<0.0001

Fonte: elaboração própria a partir dos dados da PNAD.

Tabela 11 – Estimativas para o modelo de regressão bivariada do rendimento e horas trabalhadas com a distribuição normal bivariada.

Variável dependente	Variável independente	Estimativa	Erro Padrão	Valor W	P-valor
Salário	σ_1	0.546	9e-04	577.839	<0.0001
	σ_2	0.5091	9e-04	579.5508	<0.0001
	(Intercept)	5.6254	0.0185	303.8767	<0.0001
	Sexo	0.2975	0.003	99.7119	<0.0001
	Idade	0.0514	7e-04	68.683	<0.0001
	Idade ²	-6e-04	0	-59.683	<0.0001
	Raça	0.0983	0.003	32.3605	<0.0001
	Estado Civil	0.0064	0.0068	0.9348	0.3499
	Educação	0.0467	5e-04	102.0093	<0.0001
	Pará	-0.3185	0.0062	-51.3509	<0.0001
	Ceará	-0.3502	0.0058	-59.9158	<0.0001
	Pernambuco	-0.3497	0.0056	-61.9095	<0.0001
	Bahia	-0.312	0.0057	-54.9799	<0.0001
	Minas Gerais	-0.0932	0.0054	-17.2204	<0.0001
	Rio de Janeiro	-0.0794	0.005	-15.816	<0.0001
	Paraná	0.0121	0.0065	1.8658	0.0621
	Rio Grande do Sul	-0.0884	0.0051	-17.3987	<0.0001
	Distrito Federal	0.086	0.0061	14.1187	<0.0001
	2014	0.0024	0.0033	0.722	0.4703
	2015	-0.059	0.0033	-17.877	<0.0001
	Alta	0.886	0.0079	112.0996	<0.0001
	Média-alta	0.3708	0.0072	51.7782	<0.0001
	Média	0.186	0.0068	27.482	<0.0001
	Média-baixa	0.05	0.0066	7.5617	<0.0001
	Agrícola	-0.4305	0.0163	-26.407	<0.0001
	Atividades industriais	-0.2038	0.0082	-24.8468	<0.0001
	Construção	-0.0783	0.0088	-8.8668	<0.0001
	Comércio, alimentação e outros	-0.211	0.0077	-27.2908	<0.0001
	Educação, saúde e serviços sociais	-0.2547	0.0073	-34.8405	<0.0001
	Outros serviços	-0.2144	0.0077	-27.9206	<0.0001
	Sem-Carteira	-0.1955	0.004	-49.0033	<0.0001
	Conta-Própria	-0.1363	0.0038	-35.9669	<0.0001
	Funcionário-Público	0.24	0.0068	35.433	<0.0001
	Sindicato	0.1223	0.0037	32.8447	<0.0001
	Previdência	0.3526	0.0065	54.3776	<0.0001
	Experiência	0.0122	2e-04	61.6307	<0.0001
Horas trabalhadas	(Intercept)	3.2736	0.0173	189.1109	<0.0001
	Sexo	0.0836	0.0028	29.7048	<0.0001
	Idade	0.0131	7e-04	18.421	<0.0001
	Idade ²	-2e-04	0	-18.2447	<0.0001
	Raça	-1e-04	0.0028	-0.0311	0.9752
	Estado Civil	-0.0084	0.0064	-1.3255	0.185
	Educação	-9e-04	4e-04	-2.1805	0.0292
	Pará	0.0025	0.0058	0.4379	0.6614
	Ceará	0.0171	0.0054	3.137	0.0017
	Pernambuco	-0.0154	0.0053	-2.9334	0.0034
	Bahia	-0.0444	0.0053	-8.3842	<0.0001
	Minas Gerais	0.0117	0.005	2.3154	0.0206
	Rio de Janeiro	-0.0873	0.0047	-18.6599	<0.0001
	Paraná	-0.0125	0.006	-2.0587	0.0395
	Rio Grande do Sul	0.0176	0.0047	3.7156	2e-04
	Distrito Federal	-0.0273	0.0057	-4.7988	<0.0001
	2014	0.0116	0.003	3.8156	1e-04
	2015	-0.04	0.0031	-12.9923	<0.0001
	Alta	0.1231	0.0073	16.8078	<0.0001
	Média-alta	0.099	0.0067	14.8496	<0.0001
	Média	0.1455	0.0063	23.0917	<0.0001
	Média-baixa	0.1542	0.0062	25.0599	<0.0001
	Agrícola	0.1566	0.0152	10.3231	<0.0001
	Atividades industriais	0.0199	0.0076	2.6067	0.0091
	Construção	0.0418	0.0082	5.0731	<0.0001
	Comércio, alimentação e outros	0.0645	0.0072	8.9631	<0.0001
	Educação, saúde e serviços socia	-0.0663	0.0068	-9.7416	<0.0001
	Outros serviços	-0.0151	0.0072	-2.1166	0.0343
	Sem-Carteira	-0.1635	0.0037	-44.3575	<0.0001
	Conta-Própria	-0.1744	0.0034	-50.7675	<0.0001
	Funcionário-Público	-0.0406	0.0062	-6.4987	<0.0001
	Chefe	0.0081	0.0027	3.0387	0.0024
	Menores	-0.0042	0.0027	-1.5444	0.1225
	Inatividade	0.0028	0.0039	0.7324	0.4639
	ρ	0.1905	0.0024	80.6372	<0.0001

Fonte: elaboração própria a partir dos dados da PNAD.

Tabela 12 – Estimativas para o modelo de regressão bivariada do rendimento e horas trabalhadas com a distribuição BS bivariada.

Variável dependente	Variável independente	Estimativa	Erro Padrão	Valor W	P-valor
Salário	δ_1	5.3661	0.0186	289.2067	<0.0001
	δ_2	4.3412	0.015	288.4911	<0.0001
	(Intercept)	6.0901	0.0199	306.4087	<0.0001
	Sexo	0.3092	0.0032	97.3237	<0.0001
	Idade	0.0303	8e-04	37.7251	<0.0001
	Idade ²	-3e-04	0	-28.6244	<0.0001
	Raça	0.0941	0.0033	28.9096	<0.0001
	Estado Civil	0.0223	0.0073	3.037	0.0024
	Educação	0.0568	5e-04	118.2907	<0.0001
	Pará	-0.2965	0.0067	-44.4972	<0.0001
	Ceará	-0.343	0.0062	-55.0009	<0.0001
	Pernambuco	-0.3331	0.0061	-55.0156	<0.0001
	Bahia	-0.3307	0.006	-54.7777	<0.0001
	Minas Gerais	-0.1075	0.0058	-18.4852	<0.0001
	Rio de Janeiro	-0.0761	0.0054	-14.0876	<0.0001
	Paraná	0.0341	0.007	4.8795	<0.0001
	Rio Grande do Sul	-0.0607	0.0055	-11.0906	<0.0001
	Distrito Federal	0.0977	0.0065	15.0592	<0.0001
	2014	0.0041	0.0035	1.1724	0.241
	2015	-0.0523	0.0035	-14.7843	<0.0001
	Alta	0.836	0.0083	100.2141	<0.0001
	Média-alta	0.3102	0.0077	40.4796	<0.0001
	Média	0.1485	0.0072	20.5903	<0.0001
	Média-baixa	0.0389	0.0071	5.5022	<0.0001
	Agrícola	-0.4803	0.0165	-29.1377	<0.0001
	Atividades industriais	-0.2301	0.0087	-26.478	<0.0001
	Construção	-0.0753	0.0094	-8.0129	<0.0001
	Comércio, alimentação e outros	-0.237	0.0082	-28.9335	<0.0001
	Educação, saúde e serviços sociais	-0.27	0.0077	-34.9558	<0.0001
	Outros serviços	-0.2503	0.0081	-30.9139	<0.0001
	Sem-Carteira	-0.2097	0.0043	-49.0163	<0.0001
	Conta-Própria	-0.2053	0.0039	-52.5206	<0.0001
	Funcionário-Público	0.2204	0.0073	30.393	<0.0001
	Sindicato	0.1322	0.0041	32.6318	<0.0001
	Previdência	0.3658	0.0068	54.0879	<0.0001
	Experiência	0.0121	2e-04	58.319	<0.0001
Horas trabalhadas	(Intercept)	3.4455	0.0218	157.7246	<0.0001
	Sexo	0.0834	0.0035	23.5471	<0.0001
	Idade	0.0041	9e-04	4.5606	<0.0001
	Idade ²	0	0	-3.6685	2e-04
	Raça	-0.0064	0.0036	-1.8062	0.0709
	Estado Civil	-0.0274	0.0078	-3.5129	4e-04
	Educação	0.0054	5e-04	10.1285	<0.0001
	Pará	0.0812	0.0074	10.9039	<0.0001
	Ceará	0.0808	0.007	11.6125	<0.0001
	Pernambuco	0.0434	0.0067	6.4886	<0.0001
	Bahia	-0.0352	0.0066	-5.3399	<0.0001
	Minhas Gerais	0.1117	0.0065	17.0868	<0.0001
	Rio de Janeiro	-0.1851	0.0056	-33.2668	<0.0001
	Paraná	0.0394	0.0077	5.103	<0.0001
	Rio Grande do Sul	0.0858	0.006	14.238	<0.0001
	Distrito Federal	-0.0271	0.0071	-3.8363	1e-04
	2014	0.0602	0.0039	15.435	<0.0001
	2015	-0.0721	0.0038	-18.8028	<0.0001
	Alta	0.0487	0.0092	5.2985	<0.0001
	Média-alta	0.0276	0.0083	3.3047	0.001
	Média	0.0961	0.0079	12.2012	<0.0001
	Média-baixa	0.1139	0.0077	14.7894	<0.0001
	Agrícola	0.1503	0.0194	7.7594	<0.0001
	Atividades industriais	0.0235	0.0098	2.3959	0.0166
	Construção	0.0232	0.0105	2.2082	0.0272
	Comércio, alimentação e outros	0.0456	0.0092	4.9499	<0.0001
	Educação, saúde e serviços sociais	-0.0743	0.0087	-8.5393	<0.0001
	Outros serviços	-0.0354	0.0091	-3.8701	1e-04
	Sem-Carteira	-0.1385	0.0046	-29.8113	<0.0001
	Conta-Própria	-0.1497	0.0043	-34.9072	<0.0001
	Funcionário-Público	-0.0374	0.008	-4.6759	<0.0001
	Chefe	0.0103	0.0034	3.0466	0.0023
	Menores	0.0097	0.0034	2.8275	0.0047
	Inatividade	0.0028	0.0049	0.5835	0.5595
	ρ	0.1641	0.0025	64.5008	<0.0001

Fonte: elaboração própria a partir dos dados da PNAD.

B Código em R

```
#####
## Pacotes
#####

#install.packages("readstata13")
#devtools::install_github("santosneto/RBS")
library(readstata13)
library(stargazer)
library(xtable)
library(devtools)
library(RBS)
library(numDeriv)
library(heavy)

#####
## Selecionando caminho
#####

## Lendo as funcoes usadas para estimacao

source("funcoes_regressao_bivariada_normal.R")
source("funcoes_regressao_bivariada_t.R")
source("funcoes_regressao_bivariada_bs.R")
source("ellipticalR/elliptical.r")
## Selecionar o caminho para leitura dos dados
setwd("stata_codigos")
#####
## Lendo dados stata
#####

dat <- read.dta13("filtrofinal.dta")

head(dat)

ano <- dat$ano
dat <- dat[ano=="2013",]

head(dat)
dim(dat)
```

```

sapply(dat,class)
sapply(dat,mode)

##### variaveis dependentes #####

y <- dat$y    ## salario-hora
hist (y)

r <- dat$r    ## salario
hist (r)

h <- dat$h    ##horas
hist (h)

##### afetam ambas as variaveis #####

sexo          <- dat$sexo
idade         <- dat$idade
idade2        <- dat$idade2
cor           <- dat$cor
estadocivil   <- dat$estadocivil
#
educacao      <- dat$educacao
#
uf1           <- dat$uf1    ## Referencia uf7
uf2           <- dat$uf2
uf3           <- dat$uf3
uf4           <- dat$uf4
uf5           <- dat$uf5
uf6           <- dat$uf6
uf8           <- dat$uf8
uf9           <- dat$uf9
uf10          <- dat$uf10
#
ano2          <- dat$ano2    ## Referencia ano1
ano3          <- dat$ano3
#
categ1        <- dat$categ1  ## Referencia categ5
categ2        <- dat$categ2  ## ocupacao
categ3        <- dat$categ3
categ4        <- dat$categ4
#
setor1        <- dat$setor1   ## Referencia setor5
setor2        <- dat$setor2   ## Grupamento de atividade

```

```

setor3      <- dat$setor3
setor4      <- dat$setor4
setor6      <- dat$setor6
setor7      <- dat$setor7
#
#
contrato2    <- dat$contrato2  ## Referencia contrato1
contrato3    <- dat$contrato3  ## Posicao de ocupacao
contrato4    <- dat$contrato4

##### afetam oferta de trabalho e nao ofertam salario #####
## h
chefe        <- dat$chefe
menores      <- dat$menores
desemprego   <- dat$desemprego

# sexo + idade + idade2 + cor + estadocivil + educacao + uf1 + uf2 +
#uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
# ano2 + ano3 + categ1 + categ2 + categ3 + categ4 + setor1 + setor2 +
#setor3 + setor4 + setor6 + setor7 +
# contrato2 + contrato3 + contrato4 +
# chefe + menores + desemprego

##### afetam salario nao afetam oferta #####
## r
sindicato    <- dat$sindicato
previdencia  <- dat$previdencia
experencia   <- dat$experencia

# sexo + idade + idade2 + cor + estadocivil + educacao + uf1 + uf2 + uf3 +
#uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
# ano2 + ano3 + categ1 + categ2 + categ3 + categ4 + setor1 + setor2 +
#setor3 + setor4 + setor6 + setor7 +
# contrato2 + contrato3 + contrato4 +
# sindicato + previdencia + experencia

#####
## Regressoes separadas - normal
#####

## regressao do salario-hora
# fit_y_normal <- glm(y ~ sexo + cor + estadocivil + chefe + idade +
#idade2 + educacao + categ1 + categ2 + categ3 + categ4 + sindicato +
# previdencia + contrato2 + contrato3 + contrato4 + uf1 + uf2 + uf3 + uf4 +
#uf5 + uf6 + uf8 + uf9 + uf10 + dano2 + dano3,

```

```

# family = gaussian(link = "identity"))
#
# summary(fit_y_normal)
# #newobject<-xtable(fit_y_normal)
# #print.xtable(newobject, type="latex", file="filename.tex")
# AIC(fit_y_normal)
# BIC(fit_y_normal)

## regressao do salario
fit_r_normal <- glm(r ~ sexo + idade + idade2 + cor + estadocivil +
#educacao + uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
ano2 + ano3 + categ1 + categ2 + categ3 + categ4 +
#setor1 + setor2 + setor3 + setor4 + setor6 + setor7 +
contrato2 + contrato3 + contrato4 +
sindicato + previdencia + experiencia,family =
gaussian(link = "identity"))

summary(fit_r_normal)
AIC(fit_r_normal)
BIC(fit_r_normal)

##regressao das horas
fit_h_normal <- glm (h ~ sexo + idade + idade2 + cor + estadocivil +
educacao + uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
ano2 + ano3 + categ1 + categ2 + categ3 +
categ4 + setor1 + setor2 + setor3 + setor4 +
setor6 + setor7 +
contrato2 + contrato3 + contrato4 +
chefe + menores + desemprego ,
family = gaussian(link = "identity"))

summary(fit_h_normal)
AIC(fit_h_normal)
BIC(fit_h_normal)

# newobject<-xtable(fit_h_normal)
# print.xtable(newobject, type="latex", file="filename.tex")

#####
## Regressoes separadas - t
#####

# ## regressao do salario-hora
#

```

```

# fit_y_t <-elliptical(formula=y ~ sexo + cor + estadocivil + chefe +
idade + idade2 + educacao + categ1 + categ2 + categ3 + categ4 + sindicato +
#
#               previdencia + contrato2 + contrato3 + contrato4 +
uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 + dano2 + dano3,
#
#               family=Student(4))
# summary(fit_y_t)
#
#
# #newobject<-xtable(fit_y_t)
# #print.xtable(newobject, type="latex", file="filename.tex")

## regressao do salario
fit_r_t      <- elliptical(formula=r ~ sexo + idade + idade2 + cor +
estadocivil + educacao +  uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
#
#               ano2 + ano3 +  categ1 + categ2 + categ3 +
categ4  + setor1 + setor2 + setor3 + setor4 +
setor6 + setor7 +
#
#               contrato2 + contrato3 + contrato4 +
sindicato + previdencia + experiencia,
#
#               family=Student(4))
summary(fit_r_t)

AIC_ <- 255491 + 2*37
BIC_ <- 255491 + 37*log(length(r))

##regressao das horas
fit_h_t      <- elliptical(formula=h ~ sexo + idade + idade2 + cor +
estadocivil + educacao +  uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
#
#               ano2 + ano3 +  categ1 + categ2 + categ3 + categ4  +
setor1 + setor2 + setor3 + setor4 + setor6 + setor7 +
#
#               contrato2 + contrato3 + contrato4 +
chefe + menores + desemprego ,
#
#               family=Student(4))
summary(fit_h_t)

# newobject<-xtable(fit_h_t)
# print.xtable(newobject, type="latex", file="filename.tex")

#####
## Regressoes separadas - Birnbaum-Saunders
#####

```

```

# ## regressao do salario-hora
# fit_y_bs      <- gamlss(exp(y) ~  sexo + cor + estadocivil + chefe +
idade + idade2 + educacao + categ1 + categ2 + categ3 + categ4 + sindicato +
#
#               previdencia + contrato2 + contrato3 + contrato4 +
uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 +
#
#               uf10 + dano2 + dano3, family=RBS(mu.link="log"),
method=CG() )
# resy <- summary(fit_y_bs)

## regressao do salario
fit_r_bs      <- gamlss (exp(r) ~ sexo + idade + idade2 + cor +
estadocivil + educacao +  uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 +
uf10 +
#
#               ano2 + ano3 +  categ1 + categ2 + categ3 + categ4  +
setor1 + setor2 + setor3 + setor4 + setor6 + setor7 +
contrato2 + contrato3 + contrato4 +
sindicato + previdencia + experiencia ,
family=RBS(mu.link="log"),method=CG() )
resy <- summary(fit_r_bs)

##regressao das horas
fit_h_bs      <- gamlss (exp(h) ~  sexo + idade + idade2 + cor + estadocivil +
educacao +  uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
#
#               ano2 + ano3 +  categ1 + categ2 + categ3 + categ4  +
setor1 + setor2 + setor3 + setor4 + setor6 + setor7 +
contrato2 + contrato3 + contrato4 +
chefe + menores + desemprego ,
family=RBS(mu.link="log"),method=CG() )
resy <- summary(fit_h_bs)

#AIC(fit_h_bs)
#BIC(fit_h_bs)

#resy <- newobject<-xtable(cbind(rownames(resy),(round(resy,4))))
#print.xtable(newobject, type="latex", file="filename.tex")

#####
## Regressoes bivariadas usando a distribuicao normal
#####

r <- dat$r    ## salario

```



```

h <- dat$h      ##horas

Zx1      <- model.matrix( ~ sexo + idade + idade2 + cor + estadocivil +
educacao +  uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
                        ano2 + ano3 +  categ1 + categ2 + categ3 +
                        categ4  + setor1 + setor2 + setor3 + setor4 +
                        setor6 + setor7 +
                        contrato2 + contrato3 + contrato4 +
                        sindicato + previdencia + experiencia )
Zx2      <- model.matrix( ~ sexo + idade + idade2 + cor + estadocivil +
educacao +  uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
                        ano2 + ano3 +  categ1 + categ2 + categ3 +
                        categ4  + setor1 + setor2 + setor3 +
                        setor4 + setor6 + setor7 +
                        contrato2 + contrato3 + contrato4 +
                        chefe + menores + desemprego )

pz1      <- ncol(Zx1)
pz2      <- ncol(Zx2)

Estimates_Bivariate_Normal <- profileLokligNormal(t1=r,t2=h,
Xz1=Zx1,Xz2=Zx2,pz1=pz1,pz2=pz2)

## Estimates
sigma1_Normal <- Estimates_Bivariate_Normal$sigma1Estimate
sigma2_Normal <- Estimates_Bivariate_Normal$sigma2Estimate
betas1_Normal <- Estimates_Bivariate_Normal$betas1Estimate
betas2_Normal <- Estimates_Bivariate_Normal$betas2Estimate
rho_Normal    <- Estimates_Bivariate_Normal$rhoEstimate

round(c(sigma1_Normal,sigma2_Normal,betas1_Normal,betas2_Normal,rho_Normal),3)

## Estimativas, erros-padroes e p-valores
valueH      <- -hessian(func=loglikNormalSE,x=c(sigma1_Normal,sigma2_Normal,
betas1_Normal,betas2_Normal,rho_Normal),t1=r,t2=h,Xz1=Zx1,Xz2=Zx2,pz1=pz1,pz2=pz2)
Hess        <- solve(valueH)
SE          <- sqrt(diag(Hess));SE
se.coef     = SE
tval        = c(sigma1_Normal,sigma2_Normal,betas1_Normal,betas2_Normal,rho_Normal)/se.coef
matcoef     = cbind(c(sigma1_Normal,sigma2_Normal,betas1_Normal,
betas2_Normal,rho_Normal), se.coef, tval, 2*(1-pnorm(abs(tval))))
dimnames(matcoef) = list(names(tval), c(" Estimate",
                                         " Std. Error", " t value", "Pr(>|t|)"))

cat("
nCoefficient(s):

```

```

n")
printCoefmat(matcoef, digits = 6, signif.stars = TRUE)

#newobject<-xtable(cbind(c("sigma1","sigma2",colnames(Zx1),colnames(Zx2),
"rho"),round(matcoef),4) ))
#print.xtable(newobject, type="latex", file="filename.tex")

## information criteria
logNormal <- loglikNormalSE(par=c(sigma1_Normal,sigma2_Normal,betas1_Normal,
betas2_Normal,rho_Normal),t1=r,t2=h,Xz1=Zx1,Xz2=Zx2,pz1=pz1,pz2=pz2)
AICNormal <- -2*logNormal + 2*length(c(sigma1_Normal,sigma2_Normal,
betas1_Normal,betas2_Normal,rho_Normal))
BICNormal <- -2*logNormal + log(length(r))*length(c(sigma1_Normal,
sigma2_Normal,betas1_Normal,betas2_Normal,rho_Normal))
AICNormal
BICNormal

#####
## Regressoes bivariadas usando a distribuicao t
#####

r <- dat$r ## salario
h <- dat$h ##horas

Zx1 <- model.matrix( ~ sexo + idade + idade2 + cor + estadocivil +
educacao + uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
ano2 + ano3 + categ1 + categ2 + categ3 +
categ4 + setor1 + setor2 + setor3 + setor4 +
setor6 + setor7 +
contrato2 + contrato3 + contrato4 +
sindicato + previdencia + experiencia )
Zx2 <- model.matrix( ~ sexo + idade + idade2 + cor + estadocivil +
educacao + uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
ano2 + ano3 + categ1 + categ2 + categ3 +
categ4 + setor1 + setor2 + setor3 + setor4 +
setor6 + setor7 +
contrato2 + contrato3 + contrato4 +
chefe + menores + desemprego )

pz1 <- ncol(Zx1)
pz2 <- ncol(Zx2)

Estimates_Bivariate_t <- profileLokligt(t1=r,t2=h,nu=4,

```

```

Xz1=Zx1,Xz2=Zx2,pz1=pz1,pz2=pz2)

## Estimates
sigma1_t <- Estimates_Bivariate_t$sigma1Estimate
sigma2_t <- Estimates_Bivariate_t$sigma2Estimate
betas1_t <- Estimates_Bivariate_t$betas1Estimate
betas2_t <- Estimates_Bivariate_t$betas2Estimate
rho_t    <- Estimates_Bivariate_t$rhoEstimate

round(c(sigma1_t,sigma2_t,betas1_t,betas2_t,rho_t),3)

numDeriv::hessian
## Estimativas, erros-padroes e p-valores
valueH    <- - numDeriv::hessian(func=logliktSE,x=c(sigma1_t,sigma2_t,betas1_t,
betas2_t,rho_t),t1=r,t2=h,nu=4,Xz1=Zx1,Xz2=Zx2,pz1=pz1,pz2=pz2)
Hess      <- solve(valueH)
SE        <- sqrt(diag(Hess));SE
se.coef = SE
tval = c(sigma1_t,sigma2_t,betas1_t,betas2_t,rho_t)/se.coef
matcoef = cbind(c(sigma1_t,sigma2_t,betas1_t,betas2_t,rho_t), se.coef,
tval, 2*(1-pnorm(abs(tval))))
dimnames(matcoef) = list(names(tval), c(" Estimate",
                                         " Std. Error", " t value", "Pr(>|t|)"))

cat("
nCoefficient(s):
n")
printCoefmat(matcoef, digits = 6, signif.stars = TRUE)

#newobject<-xtable(cbind(c("sigma1","sigma2",colnames(Zx1),colnames(Zx2),
"rho"),round((matcoef),4) ))
#print.xtable(newobject, type="latex", file="filename.tex")

## information criteria
logt <- logliktSE(par=c(sigma1_t,sigma2_t,betas1_t,betas2_t,rho_t),t1=r,t2=h,
nu=4,Xz1=Zx1,Xz2=Zx2,pz1=pz1,pz2=pz2)
AICt <- -2*logt+ 2*length(c(sigma1_t,sigma2_t,betas1_t,betas2_t,rho_t))
BICt <- -2*logt+ log(length(r))*length(c(sigma1_t,sigma2_t,betas1_t,
betas2_t,rho_t))
AICt
BICt

#####
## Regressoes bivariadas usando a distribuicao BS
#####

```

```

r <- exp(dat$r)    ## salario
h <- exp(dat$h)    ##horas

Zx1      <- model.matrix( ~ sexo + idade + idade2 + cor + estadocivil +
educacao +  uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
                        ano2 + ano3 +  categ1 + categ2 + categ3 +
                        categ4  + setor1 + setor2 + setor3 + setor4 +
                        setor6 + setor7 +
                        contrato2 + contrato3 + contrato4 +
                        sindicato + previdencia + experiencia )
Zx2      <- model.matrix( ~ sexo + idade + idade2 + cor + estadocivil +
educacao +  uf1 + uf2 + uf3 + uf4 + uf5 + uf6 + uf8 + uf9 + uf10 +
                        ano2 + ano3 +  categ1 + categ2 + categ3 +
                        categ4  + setor1 + setor2 + setor3 + setor4 +
                        setor6 + setor7 +
                        contrato2 + contrato3 + contrato4 +
                        chefe + menores + desemprego )

pz1      <- ncol(Zx1)
pz2      <- ncol(Zx2)

Estimates_Bivariate_BS <- profileLokligBS(t1=r,t2=h,Xz1=Zx1,Xz2=Zx2,pz1=pz1,
pz2=pz2,betas1guess=as.vector(betas1_Normal),betas2guess=as.vector(betas2_Normal))

## Estimates
delta1_BS <- Estimates_Bivariate_BS$delta1Estimate
delta2_BS <- Estimates_Bivariate_BS$delta2Estimate
betas1_BS <- Estimates_Bivariate_BS$betas1Estimate
betas2_BS <- Estimates_Bivariate_BS$betas2Estimate
rho_BS   <- Estimates_Bivariate_BS$rhoEstimate

round(c(delta1_BS,delta2_BS,betas1_BS,betas2_BS,rho_BS),3)

## Estimativas, erros-padroes e p-valores
valueH   <- -hessian(func=loglikBSSE,x=c(delta1_BS,delta2_BS,betas1_BS,
betas2_BS,rho_BS),t1=r,t2=h,Xz1=Zx1,Xz2=Zx2,pz1=pz1,pz2=pz2)
Hess     <- solve(valueH)
di       <- diag(Hess)
for(i in 1:length(di)) if(di[i]<0) di[i] <- di[i]*(-1)
SE       <- sqrt(di);SE
se.coef  = SE
tval     = c(delta1_BS,delta2_BS,betas1_BS,betas2_BS,rho_BS)/se.coef
matcoef  = cbind( c(delta1_BS,delta2_BS,betas1_BS,betas2_BS,rho_BS),
se.coef, tval, 2*(1-pnorm(abs(tval))))
dimnames(matcoef) = list(c("delta1","delta2",colnames(Zx1),
colnames(Zx1),"rho"), c( " Estimate",
                        " Std. Error", " t value", "Pr(>|t|)"))

```

```

cat("
nCoefficient(s):
n")
printCoefmat(matcoef, digits = 6, signif.stars = TRUE)

newobject<-xtable(cbind(c("delta1","delta2",colnames(Zx1),colnames(Zx2),
"rho"),round((matcoef),4) ))
print.xtable(newobject, type="latex", file="filename.tex")

## information criteria
logbs <- loglikBSSE(par=c(delta1_BS,delta2_BS,betas1_BS,betas2_BS,rho_BS),
t1=r,t2=h,Xz1=Zx1,Xz2=Zx2,pz1=pz1,pz2=pz2)
AICbs <- -2*logbs+ 2*length(c(delta1_BS,delta2_BS,betas1_BS,betas2_BS,rho_BS))
BICbs <- -2*logbs+ log(length(r))*length(c(delta1_BS,delta2_BS,betas1_BS,
betas2_BS,rho_BS))
AICbs
BICbs

```

Referências

- BALAKRISHNAN, N.; LAI, C.-D. *Continuous Bivariate Distributions*. New York: Springer, 2009. 13, 22
- BARROS, M.; PAULA, G.; LEIVA, V. A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Analysis*, v. 14, p. 316–332, 2008. 26
- BARROS, R.; RAMOS, L.; SANTOS, E. Gender differences in brazilian labor markets. In: SCHULTZ, P. (Ed.). *Investment in women's human capital*. 1. ed. [S.l.]: University of Chicago Press, 1995. 19
- BARROS, R. P.; MENDONÇA, R. S. P.; DUARTE, R. P. N. Bem-estar, pobreza e desigualdade de renda: uma avaliação da evolução histórica e das disparidades regionais. *Texto para discussão n. 454 IPEA*, 1997. 17, 19
- BECKER, G. *Human Capital: A Theoretical and Empirical Analysis*. New York: Columbia University Press, 1975. 16
- BECKER, G. *Human capital a theoretical and empirical analysis, with special reference to education*. New York: NBER, 1993. 12, 15
- BIRNBAUM, Z.; SAUNDERS, S. A new family of life distributions. *Journal of Applied Probability*, v. 6, p. 319–327, 1969. 13
- BORJAS, G. J. *Economia do trabalho*. 5. ed. São Paulo: Artmed, 2012. 32
- BRANCO, R. C. Crescimento acelerado e o mercado de trabalho: a experiência brasileira. *Revista Brasileira de Economia*, v. 33, p. 247–285, 1979. 17, 18
- BUCHINSKY, M. Quantile regression with sample selection: estimating women's return to education in the u.s. *Empirical Economics*, v. 26, p. 87–113, 2001. 17
- CAIN, G. The economic analysis of labor market discrimination: a survey. In: ASHENFELTER, O.; LAYARD, R. (Ed.). *Handbook of Labor Economic*. 1. ed. [S.l.]: North-Holland, 1986. 16
- CAVALIERI, C. H.; FERNANDES, R. Diferenciais por gênero e cor: uma comparação entre as regiões metropolitanas brasileiras. *Revista de Economia Política*, v. 18, 1998. 45
- CHATTERJEE, S.; PRICE, B. *Regression Analysis by Example*. New York: John Wiley, 1991. 17

- CRESPO, A.; REIS, M. Sheepskin effects and relationship between earnings and education: analyzing their evolution over time in brazil. *Revista Brasileira de Economia*, v. 63, p. 209–231, 2009. 15
- CUNHA, J.; CORNACHIONE-JUNIOR, E.; MARTINS, G. Doutores em ciências contábeis: Análise sob a Óptica da teoria do capital humano. *Revista de Administração Contemporânea*, v. 14, p. 532–557, 2010. 15
- CURI, A. Z. G.; MENEZES-FILHO, N. A. Os determinantes das transições ocupacionais no mercado de trabalho brasileiro. *Anais do XXXII Encontro Nacional da Anpec*, 2004. 16
- FERNANDES, R. Desigualdade salarial: aspectos teóricos. In: CORSEUIL, C. H. e. a. (Ed.). *Estrutura salarial: aspectos conceituais e novos resultados para o Brasil*. Rio de Janeiro: IPEA, 2002. 46
- GAREN, J. The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica*, v. 52, p. 1199–1218, 1984. 16, 17
- GONZAGA, G.; LEITE, P. G. G. P.; MACHADO, D. C. Quem trabalha muito e quem trabalha pouco no brasil? In: *Anais do XIII Encontro Nacional De Estudos Populacionais*. Ouro Preto: ABEP, 2002. 45, 46
- HECKMAN, J. Shadow prices, market wages and labor supply. *Econometrica*, v. 42, p. 679–694, 1974. 17
- HECKMAN, J. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, v. 5, p. 475–492, 1976. 17, 21
- HECKMAN, J. Sample selection bias as a specification error. *Econometrica*, v. 47, p. 153–161, 1979. 17, 51
- HECKMAN, J.; TOBIAS, J.; VYTLACIL, E. Simple estimators for treatment parameters in a latent variable framework with an application to estimating the returns to schooling. *NBER Working Paper*, v. 7950, 2000. 12, 15
- IBACACHE-PULGAR, G.; PAULA, G.; GALEA, M. On influence diagnostics in elliptical multivariate regression models with equicorrelated random errors. *Statistical Modelling*, v. 16, p. 14–21, 2014. 40
- JANNUZZI, P. d. M. Status socioeconômico das ocupações brasileiras: medidas aproximativas para 1980, 1991 e anos 90. *Revista Brasileira de Estatística*, v. 2, p. 47–74, 2001. 32

- JOHNSON, N.; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Distributions*. New York, US: Wiley, 1995. v. 2. 13, 22
- JUNIOR, E. R. S.; MENEZES-FILHO, N. A.; FERREIRA, P. C. Migração, seleção e diferenças regionais de renda no brasil. *Pesquisa e Planejamento Econômico*, v. 35, p. 299–332, 2005. 19
- LANGONI, C. *Distribuição da renda e desenvolvimento econômico do Brasil*. Rio de Janeiro: Expressão e Cultura, 1973. 17
- LAU, L. J. et al. Education and economic growth some cross-sectional evidence from brazil. *Journal of Development Economics*, v. 41, p. 45–70, 1993. 46
- LEIVA, V. et al. Birnbaum-Saunders regression model: a new approach. *Statistical Modelling*, v. 14, p. 21–48, 2014. 27
- MACIEL, M.; CAMPÊLO, A.; RAPOSO, M. A dinâmica das mudanças na distribuição salarial e no retorno em educação para mulheres: uma aplicação de regressão quantílica. In: *Anais do XXIX Encontro Nacional de Economia*. Salvador: ANPEC, 2001. 18
- MARCHANT, C.; LEIVA, V.; CYSNEIROS, F. J. A. A multivariate log-linear model for Birnbaum-Saunders distributions. *IEEE Transactions on Reliability*, v. 65, p. 816–827, 2016. 12, 39
- MATOS, R. S.; MACHADO, A. F. Diferencial de rendimento por cor e sexo no brasil (1987-2001). *Revista econômica*, v. 8, 2006. 17
- MINCER, J. Investment in human capital and personal income distribution. *Journal of Political Economy*, v. 66, p. 281–302, 1958. 15, 16
- MINCER, J. *Schooling, Experience, and Earnings*. National Bureau of Economic Research, Inc, 1974. Disponível em: <<http://EconPapers.repec.org/RePEc:nbr:nberbk:minc74-1>>. 16, 17
- MITTELHAMMER, R. C.; JUDGE, G. G.; MILLER, D. J. *Econometric Foundations*. New York, US: Cambridge University Press, 2000. 21
- MONSUETO, S. E. *Distribuição de Renda no Brasil: A Situação dos 25% Mais Pobres*. Dissertação (Dissertação de Mestrado) — Universidade Federal de Minas Gerais, Belo Horizonte, Brasil, 2003. 18
- OLIVEIRA, A. M. H. C. Indicadores da segregação ocupacional por sexo no brasil. In: *Anais do XI Encontro Nacional De Estudos Populacionais*. [S.l.]: ABEP, 1998. 17

- REIS, M.; RAMOS, L. Escolaridade dos pais, desempenho no mercado de trabalho e desigualdade de rendimentos. *Revista Brasileira de Economia*, v. 65, p. 177–205, 2011. 15
- RIECK, J.; NEDELMAN, J. A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, v. 3, p. 51–60, 1991. 29
- SACHCIDA, A.; LOUREIRO, P. R. A.; MENDONÇA, M. Um estudo sobre retorno em escolaridade no brasil. *Revista Brasileira de Economia*, v. 58, p. 249–265, 2004. 16
- SANTOS-NETO, M. et al. On new parameterizations of the Birnbaum-Saunders distribution. *Pakistan Journal of Statistics*, v. 28, p. 1–26, 2012. 13, 14, 27
- SAULO, H. et al. On a bivariate Birnbaum-Saunders distribution parameterized by its means. *Working Paper*, 2016. 13, 22, 28
- SCHULTZ, T. *O capital humano: investimentos em educação e pesquisa*. Rio de Janeiro: Zahar Editores, 1973. 15
- SEDLACEK, G.; SANTOS, E. A mulher cônjuge no mercado de trabalho como estratégia de geração da renda familiar. *Pesquisa e Planejamento Econômico*, v. 21, p. 449–470, 1991. 19
- SENNA, J. Escolaridade, experiência no trabalho e salários no brasil. *Revista Brasileira de Economia*, v. 30, p. 163–193, 1976. 17
- STAPLETON, J. *Linear Statistical Models*. New York: John Wiley, 1995. 17
- TANNURI-PIANTO, M.; PIANTO, D. A choice at the top and segmentation at the bottom: A quantile regression approach. In: *Anais do XXIV Encontro Brasileiro de Econometria*. [S.l.]: ANPEC, 2002. 16
- VAZ, D. V.; HOFFMANN, R. Segregação ocupacional por sexo no setor público brasileiro no período 1995 e 2008. *Revista ABET*, v. 10, 2011. 17
- VILCA, F.; BALAKRISHNAN, N.; ZELLER, C. The bivariate sinh-elliptical distribution with applications to Birnbaum-Saunders distribution and associated regression and measurement error models. *Computational Statistics and Data Analysis*, v. 80, p. 1–16, 2014. 40
- WEISBERG, S. *Applied Linear Regression*. 4. ed. New Jersey: Wiley, 2013. 35
- WILLIS, R.; ROSEN, S. Education and self selection. *Journal of Political*, v. 87, p. 7–36, 1979. 16