



Universidade Federal da Paraíba  
Centro de Ciências Sociais Aplicadas  
Programa de Pós-Graduação em Economia

# Migração seletiva de retorno e distribuição salarial: Evidências para população migrante em São Paulo

JULIANE DO CARMO DUARTE MAGALHÃES

João Pessoa - PB  
2020

# **Migração seletiva de retorno e distribuição salarial: Evidências para população migrante em São Paulo**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba - UFPB, em cumprimento às exigências do Curso de Mestrado em Economia.

Universidade Federal da Paraíba  
Centro de Ciências Sociais Aplicadas  
Programa de Pós-Graduação em Economia

Orientador: Dr. Hilton Martins de Brito Ramalho  
Coorientador: Dr. Alessio Tony Cavalcanti de Almeida

João Pessoa - PB

2020

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

M188m Magalhaes, Juliane do Carmo Duarte.

Migração seletiva de retorno e distribuição salarial:  
evidências para população migrante em São Paulo /  
Juliane do Carmo Duarte Magalhaes. - João Pessoa, 2020.  
75 f. : il.

Orientação: Hilton Martins de Brito Ramalho.

Coorientação: Aléssio Tony Cavalcanti de Almeida.

Dissertação (Mestrado) - UFPB/CCSA.

1. Migração de Retorno. 2. Economia. 3. Desigualdade  
Salarial. 4. Autosseleção - Migração. I. Ramalho,  
Hilton Martins de Brito. II. de Almeida, Aléssio Tony  
Cavalcanti. III. Título.

UFPB/BC

CDU 331.556.2

Dedico essa dissertação aos meus pais, Francisco e Maria Lúcia, em gratidão ao seu esforço e dedicação e pela força que me ensinaram a ter para lutar todos os dias pelos meus sonhos.

# Agradecimentos

Ao Senhor Deus que me sustenta desde muito cedo nesta vida, sem o qual eu não teria chegado ao fim desta jornada.

Ao professor Hilton Martins de Brito Ramalho pela dedicação, paciência, organização e excelente orientação ao longo desse processo.

Ao meu amado esposo Adaicles, meu maior incentivador, pelo seu apoio incondicional, pela companhia e compreensão quando da minha ausência.

Agradeço à professora Ana Hermínia pelo apoio no início de nossa jornada no estado da Paraíba, pelas palavras de incentivo e pelos ensinamentos na graduação.

A todos que, de alguma forma, contribuíram para realização desse trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

# Resumo

Essa dissertação teve como objetivo principal analisar os efeitos da migração de retorno sobre a distribuição salarial dos migrantes interestaduais residentes em São Paulo. Para recuperar a distribuição da população total migrante na presença de migração seletiva de retorno foram utilizados os dados do Censo Demográfico e 2010 e um método de estimador que leva em conta a seleção em inobserváveis. Os achados deste estudo apontam para seleção negativa em observáveis e não observáveis no fluxo de partida para o estado de São Paulo. Tendo em vista que migrantes retornados têm características produtivas não observadas melhores quando comparadas aos migrantes permanentes. A migração de retorno piora a desigualdade dentro da população de migrantes migrantes no destino devido à saída de trabalhadores positivamente selecionados. A lacuna entre não naturais e nativos, por sua vez, seria parcialmente reduzida se todos permanecessem no estado.

**Palavras-chave:** migração de retorno; autosseleção; desigualdade salarial.

# Abstract

The main objective of this dissertation was to analyze the effects of return migration on the wage distribution of interstate migrants residing in São Paulo. To recover the distribution of the total migrant population in the presence of selective return migration, data from the 2010 Demographic Census and an estimator method that takes into account the selection in unobservable were used. The findings of this study point to a negative selection in observables and non-observables in the outflow to the state of São Paulo. Bearing in mind that returned migrants have productive characteristics not observed better when compared to permanent migrants. Return migration worsens inequality outmigration of positively selected workers. The gap between migrants and natives, in turn, would be partially reduced if all migrants remained in the state

**Keywords:** return migration, self-selection, wage inequality.

# Lista de tabelas

Tabela 1 – Distribuição dos migrantes interestaduais segundo estado de residência na data censitária e estado de residência anterior - 2010 . . . .	24
Tabela 2 – Distribuição dos migrantes permanentes em São Paulo segundo faixa de instrução e por região de residência anterior (nascimento) . . . .	26
Tabela 3 – Médias das características socioeconômicas e demográficas observáveis de migrantes e não migrantes (Amostra) . . . . .	37
Tabela 4 – Estimativas das equações de seleção e salários . . . . .	40
Tabela 5 – Características médias dos migrantes do GAP versus demais migrantes	44
Tabela 6 – Decis das distribuições de salários dos migrantes com e sem migração de retorno . . . . .	47
Tabela 7 – Diferenças salariais entre migrantes e naturais de São Paulo - Cenário real e contrafactual . . . . .	51
Tabela 8 – Decis de distribuições de salários dos migrantes com e sem migração de retorno por nível educacional . . . . .	52
Tabela A1 – Variáveis socioeconômicas, educacionais e demográficas que foram utilizadas nos modelos de seleção e salários . . . . .	67
Tabela A2 – Características das famílias de cópulas selecionadas . . . . .	68
Tabela A3 – Critério de informação AIC: seleção do modelo segundo a função cópula - modelo principal . . . . .	68
Tabela A4 – Critério de informação AIC: seleção do modelo segundo a função cópula - amostra de mulheres . . . . .	69
Tabela A5 – Critério de informação AIC: seleção do modelo segundo a função cópula - amostra incluindo mulheres . . . . .	70
Tabela A6 – Decis de distribuições de salários dos migrantes com e sem migração de retorno - nível de instrução superior e pós graduação . . . . .	71
Tabela A7 – Teste para restrição de exclusão (filho nascido em SP) - Huber e Mellace (2014) . . . . .	71
Tabela A8 – Equações de seleção e salários - modelos incluindo mulheres . . . .	72
Tabela A9 – Equações de seleção e salários - abordagem paramétrica . . . . .	73
Tabela A10 – Estimativas das equações de seleção e salários - modelo sem restrição de exclusão (Filho nascido em SP) . . . . .	75
Tabela A11 – Regressões salariais de não migrantes paulistas . . . . .	76



# Lista de ilustrações

Figura 1 – Distribuição de migrantes em São Paulo por estado de origem - Censos de 2000 e 2010 . . . . .	25
Figura 2 – Resultados das estimativas não paramétricas das equações de seleção e salários . . . . .	43
Figura 3 – Densidade real e contrafactual do salário-hora para migrantes e paulistas - observáveis . . . . .	48
Figura 4 – Densidade real e contrafactual do salário-hora total para migrantes e paulistas . . . . .	49
Figura 5 – Densidades reais e contrafatuais do salário-hora total para migrantes e paulistas por nível educacional . . . . .	54
Figura 6 – Densidade reais e contrafatuais do salário-hora total para migrantes e paulistas - (1) Amostra de mulheres, (2) Modelo incluindo homens e mulheres e (3) Modelo sem restrição de exclusão . . . . .	58
Figura A1 – Densidade real e contrafactual do salário-hora para migrantes e paulistas - abordagem paramétrica . . . . .	74

# Sumário

1	INTRODUÇÃO . . . . .	11
2	REVISÃO DA LITERATURA . . . . .	14
2.1	Migração na teoria econômica, autosseleção e desigualdade de renda	14
2.2	Migração de retorno e autosseleção . . . . .	17
3	MIGRAÇÃO INTERESTADUAL NO BRASIL: FATOS OBSERVA- DOS . . . . .	21
3.1	Tendências Migratórias no Brasil . . . . .	21
3.2	Caracterização da migração interestadual . . . . .	23
4	METODOLOGIA . . . . .	28
4.1	Modelo Empírico . . . . .	28
4.1.1	Estimação dos parâmetros . . . . .	29
4.1.2	Estimação de densidade contrafactual . . . . .	32
4.2	Base de Dados e Tratamento . . . . .	35
5	RESULTADOS . . . . .	39
5.1	Determinantes da decisão de remigrar e salários . . . . .	39
5.2	Perfil do Grupo de Controle . . . . .	43
5.3	Efeitos da migração de retorno sobre os salários dos imigrantes . . .	45
5.4	Efeitos da migração de retorno considerando diferentes níveis edu- cacionais . . . . .	51
5.5	Análise de Robustez . . . . .	55
5.5.1	Modelo sem restrição de exclusão e validade do instrumento . . . . .	55
5.5.2	Análise incluindo mulheres . . . . .	56
6	CONSIDERAÇÕES FINAIS . . . . .	59
	REFERÊNCIAS . . . . .	62
7	APÊNDICE . . . . .	67

# 1 Introdução

Segundo dados dos censos demográficos no Brasil, houve uma expansão da migração interestadual de retorno nas últimas décadas. Em termos absolutos, a população de remigrados subiu de 1,14 milhões entre 1995 e 2000 para 1,23 milhões no quinquênio de 2005/2010, representando 22,02% e 24,52% do total de migrantes interestaduais do país, respectivamente (IBGE, 2012).

Esse regresso, no entanto, não ocorre de forma aleatória dentro do processo migratório. Segundo Biavaschi (2016), a escolha de residência ótima de um migrante racional está condicionada à decisão inicial de migração. Borjas e Bratsberg (1996), por sua vez, argumentam que a dinâmica de retorno reforça a seleção que caracterizou o fluxo de migração inicialmente. Isso significa que se o fluxo inicial de migrantes for positivamente selecionado em termos de suas habilidades inatas, então os retornados serão aqueles com habilidades inferiores dentro do grupo inicial de partida. A volta de indivíduos negativamente selecionados, no entanto, levaria a uma redução no bem-estar da população, isso porque esses indivíduos tenderiam a ser desmotivados e menos favorecidos no mercado de trabalho (SIQUEIRA, 2006).

A hipótese de autosseleção é, portanto, uma questão central nos movimentos populacionais. Ela sugere que os migrantes não são pessoas aleatórias na população de origem isto é, esses indivíduos dispõem de características não observáveis como agressividade, empreendedorismo, ambição, propensão ao risco, entre outras (CHISWICK, 1999). Além das características pessoais, os custos de migração e os diferenciais de salários líquidos entre as regiões podem influenciar a seleção dos migrantes (SJAASTAD, 1962; CATTANEO et al., 2007; MCKENZIE; RAPOPORT, 2010).

Existem evidências de que os migrantes interestaduais no Brasil são positivamente selecionados, isto é, possuem características não observáveis favoráveis ao seu sucesso no mercado de trabalho tanto em relação aos seus conterrâneos quanto em relação aos não migrantes dos estados que os recebem (JÚNIOR et al., 2005; FREGUGLIA; PROCÓPIO, 2013; GAMA; MACHADO, 2014). Ramalho e Queiroz (2011) identificam seleção positiva dos migrantes não retornados em relação aos não migrantes em fatores não observados. O grupo de retornados, por sua vez, também registra vantagens frente aos não migrantes em termos de características inatas. A conclusão dos autores foi que a migração gera benefícios mesmo quando o indivíduo opta por retornar devido aos ganhos de capital humano. Gama e Machado (2014), por sua vez, corroboram com esse resultado, os autores identificam seleção positiva em habilidades não observáveis para os remigrados e migrantes permanentes em relação aos não migrantes, mas não

comparam relações entre permanentes e retornados.

Em se tratando do progresso do migrante no destino, [Neto e Magalhães \(2004\)](#), ao estudar a migração para o estado de São Paulo, verificam que há uma lacuna significativa entre nativos e migrantes perdurando ao longo do tempo, favorecendo o grupo de paulistas. Portanto, para os migrantes, a valoração de seus atributos é feita de forma diferenciada pelo mercado, em termos de características observáveis como educação e estado de origem. Outros estudos apontam para a mesma característica de migração. Segundo [Batista e Cacciamali \(2009\)](#) esse resultado é presente entre migrantes e não migrantes na própria região Sudeste. Os autores [Assis, Costa e Mariano \(2012\)](#), ao analisar a migração entre os estados de São Paulo e Bahia identificam, mais uma vez, um hiato salarial entre baianos e paulistas, favorável ao último grupo.

O progresso econômico do migrante no destino, no entanto, pode sofrer efeitos da saída de trabalhadores. A depender da natureza de seleção desses indivíduos a desigualdade salarial entre os grupos pode ser mitigada ou intensificada. Com isso, a assimilação do migrante no destino pode ser subestimada ou superestimada se não levarmos em conta a migração seletiva de retorno ([COHEN; HABERFELD, 2001](#); [AMBROSINI et al., 2010](#)). Isso posto, surgem duas questões relevantes a serem investigadas: como se comportaria a distribuição de salários dos migrantes interestaduais na ausência da migração de retorno? A migração de retorno é um mecanismo que facilita ou dificulta a assimilação dos migrantes na região de destino?

Portanto, o objetivo geral deste trabalho é analisar o impacto da migração interestadual de retorno na distribuição de salários dos migrantes que permaneceram na região de destino. De forma específica, procura-se: (i) analisar o tipo de seletividade em características inatas dos migrantes; (ii) investigar o efeito da migração de retorno considerando grupos por níveis educacionais; (iii) verificar diferenciais salariais dentro do grupo de migrantes permanentes e entre migrantes e naturais da região de destino mediante a ausência de migração de retorno.

Para alcance dos objetivos citados serão feitas estimativas de distribuição de salários mediante dois cenários: com e sem migração de retorno, considerando o estado de São Paulo como região de destino dos fluxos de migração. O estado de São Paulo possui o maior Produto Interno Bruto (PIB) frente as demais unidades federativas do Brasil, tem um mercado de trabalho dinâmico e competitivo, além de se caracterizar como destino principal dos migrantes brasileiros ([SANTOS, 2007](#); [OLIVEIRA; ERVATTI; O'NEILL, 2011](#); [RAMALHO; FIGUEIREDO; NETTO JÚNIOR, 2016](#)).

Nesse contexto, procura-se recuperar a distribuição de salários na ausência de migração de retorno, conforme método proposto por [Biavaschi \(2016\)](#). Esse método amplia aquele apresentado por [DiNardo, Fortin e Lemieux \(1995\)](#) e aplicado em

Chiquiar e Hanson (2005) apenas para características observáveis, ao considerar a seleção dos migrantes em características não observadas. Também parte do pressuposto de que o viés de seleção amostral desaparece para migrantes com alta probabilidade de permanência na região de destino. Esse procedimento é conhecido na literatura como identificação no infinito (CHAMBERLAIN, 1986) e foi defendido por Heckman (1990) na estimação do termo constante em modelos de seleção de amostra semiparamétrica. A seleção amostral é tratada através de um modelo de regressão de cópula. Esse método assegura flexibilidade na distribuição dos dados e eficiência do modelo paramétrico, além de lidar com relações de resposta não linear das covariadas do modelos.

Destarte, esse trabalho avança em relação à literatura nacional ao considerar o efeito da migração seletiva de retorno sobre a distribuição salarial dos migrantes remanescentes no destino. Bem como pela utilização de um método ainda não explorado que considera características não observadas, identificação no infinito e a utilização de função cópula na modelagem de dependência entre o processo salarial e de migração de retorno.

Os resultados empíricos mostram que uma análise baseada apenas em características observáveis subestima o impacto da migração de retorno sobre a distribuição salarial da população migrante. No cenário em que todos os migrantes permanecem no destino (São Paulo) há um aumento de 4,28% da média salarial. Além disso, considerando características inatas, encontramos que a migração de retorno tem um impacto maior para os níveis mais baixos de rendimento. Ademais, sem migração de retorno a lacuna existente entre migrantes e paulistas, que favorece o último grupo, tende a ser parcialmente reduzida.

Afora essa introdução, esta dissertação está organizada da seguinte forma: (ii) o capítulo 2 apresenta uma revisão da literatura especializada; (iii) o capítulo 4 resume fatos observados na migração interestadual brasileira; (iv) no capítulo 4 é feito um detalhamento do modelo empírico e uma descrição dos dados e variáveis utilizadas no estudo; (v) o capítulo 5 registra os resultados empíricos e (vi) capítulo 6 é reservado às considerações finais.

## 2 Revisão da literatura

A análise da migração interna se tornou recorrente nos estudos sobre mobilidade populacional devido à importância em termos de bem estar social e econômico tanto para os indivíduos envolvidos quanto para as regiões de destino e origem. Dessa forma, é importante apresentar uma breve revisão da literatura sobre a migração, seletividade e salários na teoria econômica. Neste capítulo, também são registradas importantes evidências empíricas sobre o processo de remigração no contexto nacional e internacional.

### 2.1 Migração na teoria econômica, autosseleção e desigualdade de renda

A migração do trabalho tem sido tratada em muitas dimensões. A abordagem mais explorada coloca tal movimento como parte dos investimentos em capital humano, juntamente com a educação e os treinamentos, como defendeu [Schultz \(1961\)](#). Um dos trabalhos seminais na abordagem do capital humano no contexto da migração foi o de [Sjaastad \(1962\)](#). Esse autor indica que a migração se relaciona com os custos de forma negativa e com os rendimentos positivamente. Além disso, esses custos não se limitam aos monetários e de oportunidade, custos psíquicos de deixar a família e a comunidade devem ser considerados.

Tendo em vista que a migração é um importante mecanismo de crescimento econômico através do capital humano adquirido pela experiência, outros estudiosos se preocuparam em investigar se esses indivíduos seriam aleatórios na população. [Chiswick \(1999\)](#) identificou algumas implicações referentes à seletividade em características produtivas (motivação, propensão ao risco, perseverança) favoráveis aos migrantes econômicos, isto é, aqueles que decidem migrar tendo em vista a taxa de retorno desse movimento. Além de reforçar a ideia de uma relação inversa entre os custos e a migração, também identificou que quanto maiores esses custos, menor a propensão de retorno e maior é a seletividade favorável da migração. A principal conclusão desse autor é que há uma tendência para seletividade positiva dos migrantes econômicos, que talvez não seja tão intensa em relação aqueles que migram por motivos não econômicos.

Existe, porém, um adendo nesse aspecto autosseleção na migração referente aos atributos que tornam um migrante mais propenso ao sucesso que se refere aos atributos observáveis e não observáveis. A autosseleção pode surgir como efeito de atributos

observáveis como riqueza e educação, mas também pode surgir de características subjacentes como ambição, habilidades e empreendedorismo. Torna-se preocupante essa questão quando tomado por exemplo o grupo de altamente qualificados, quando da sua saída externalizam-se custos prejudiciais ao conjunto econômico de um país. A fuga de cérebros é, portanto, a questão de grande preocupação em relação à migração (CATTANEO et al., 2007).

Borjas (1987) apresenta a análise da seletividade a partir de um modelo de vantagens comparativas e argumenta que a migração para os Estados Unidos vai depender da distribuição dos salários do país de origem. Isto é, indivíduos com menos habilidades encontram mais incentivos em migrar do que aqueles mais qualificados, pois estes últimos são melhor recompensados em seu país de origem devido ao alto retorno da educação nesses países e alta dispersão salarial. Seguindo esse autor, Chiquiar e Hanson (2005), no entanto, encontram uma seleção positiva ou intermediária, dependendo dos custos de migração e das habilidades observadas nos migrantes e acrescentam que a migração para o exterior pode aumentar a desigualdade salarial no México por conta da migração de assalariados intermediários e médios.

De fato, o diferencial de renda nas regiões envolvidas na dinâmica migratória exerce um papel importante tanto no processo decisão, quanto nas consequências do retorno daqueles que decidiram, por razões diversas, começar uma experiência longe de casa. A teoria neoclássica é a corrente que mais destaca essa problemática, quando defende que a principal motivação dos migrantes é a diferença salarial entre as regiões que enviam e que recebem. Logo, a expectativa das pessoas é de melhorar sua situação financeira com a migração (TODARO, 1969).

As abordagens citadas consideram a decisão individual, baseada essencialmente nos ganhos econômicos dessa arbitragem. No entanto, Mincer (1978) considerou os laços familiares como importantes para decisões de migração, além de analisar o impacto dessas relações na probabilidade de migração e como essas decisões afetam salários e empregos dos membros das famílias. Nessa análise, os ganhos das famílias, ao invés dos ganhos individuais exercem força na migração. Como os laços familiares dificultam a migração, espera-se que pessoas casadas estejam menos propensas a migrar do que as solteiras. O tamanho da família também é importante, pois o retorno da migração aumenta menos que os custos com a presença de filhos e cônjuges.

Esses estudos deram as bases para que muitas evidências empíricas sobre migração, autoseleção e consequências econômicas fossem engendradas. McKenzie e Rapoport (2010), utilizando um teórico semelhante ao empregado em Chiquiar e Hanson (2005) decidiram abranger essa análise tendo em vista a série de resultados conflitantes sobre a natureza da seleção na migração México-Estados Unidos. Esse



autores destacam as redes sociais<sup>1</sup>. Com base no pressuposto de que existem custos monetários e não monetários envolvidos na decisão de migrar que variam conforme o nível educacional, redes mais fracas de migração tendem a selecionar positivamente os migrantes, já que para esse grupo os custos de migração são mais altos. No caso de comunidades fortes a seleção tende a ser negativa, pois os custos e o retorno à educação são menores.

Utilizando o modelo Roy-Borjas e considerando indivíduos heterogêneos quanto à propensão não observada de migrar, as principais conclusões de Bertoli, Moraga e Ortega (2013) para um grupo de migrantes do Equador com destinos como Estados Unidos e Espanha está relacionada às diferenças nos ganhos entre as partes. Analisando as características observáveis em romenos, Ambrosini et al. (2010) destacaram a importância da migração em relação à fuga de cérebros, indicando que a decisão de migração responde aos incentivos econômicos e indivíduos com alta qualificação tendem a migrar para onde os prêmios de salários são maiores para essa classe, como os Estados Unidos. No caso da Espanha, o prêmio salarial para quem migra é maior para os de baixa qualificação. Resumidamente, a decisão de migrar ou retornar, em termos de características observáveis, vai depender do prêmio salarial oferecido para aquela classe específica de habilidades, e isso determinará a natureza da seleção.

Há evidências da natureza de seleção no Brasil. Júnior et al. (2005) foi um dos primeiros a investigar o assunto no país. Esses autores utilizaram os dados da Pesquisa Nacional de Amostras por Domicílios (PNAD) de 1999 e uma regressão *minceriana* para o logaritmo natural do salário contra diversos controles, para testar se os migrantes interestaduais no Brasil são positivamente selecionados em relação aos não migrantes. Identificou-se que esses indivíduos não constituem um grupo aleatório na população, pois tendem a ser mais dispostos enfrentar os custos monetários e não monetários associados à migração, ou seja, características não observáveis garantem que esse grupo será favorecido em relação àqueles que não optaram pela mudança de residência. Dada essa descoberta, surge a preocupação referente aos efeitos da migração interestadual na distribuição de renda no país.

A decisão de mudar de estado pode estar relacionada aos diferenciais de renda entre as regiões ricas (principalmente o Sudeste) e as regiões mais pobres (destacando-se o Nordeste). Lima, Simões e Hermeto (2015) buscaram identificar exatamente essa questão, utilizando dados dos Censo Demográficos nos períodos de 1980-2010 e regressão logística para identificar a influência das taxas regionais de pobreza e desigualdade de renda sobre a decisão de migrar no Brasil. Foi encontrada uma relação

<sup>1</sup> Os autores seguem Massey, Goldring e Durand (1994) ao medir a magnitude da comunidade de migração. Essas redes são constituídas de indivíduos de uma determinada comunidade que já migraram. O tamanho dessas redes deve influenciar os custos de migração de um indivíduo e por conseguinte o padrão de seletividade observado (MCKENZIE; RAPOPORT, 2010).



direta entre essas variáveis, quando do sentimento de privação surge a necessidade de deslocamento. No entanto, os resultados para 2010 podem estar indicando alterações, tendo em vista o aumento dos deslocamentos de retorno à origem, especialmente Nordeste e Minas Gerais, já que esse movimento pode gerar um melhora no bem estar do remigrado e uma diminuição do sentimento de privação.

Afora outros fatores, a migração interestadual é significativa para mudanças salariais no Brasil. [Freguglia e Procópio \(2013\)](#), utilizando dados do Raismigra de 1995-2006, compararam o efeito da mudança de emprego sem migração com o caso em que há migração e mudança de firma, afim de determinar a variação do salarial do trabalhar em relação à migração propriamente dita. Tanto a mudança de firma quanto a migração interestadual são importantes fatores na determinação dos diferenciais salariais, considerando os efeitos fixos dos trabalhadores.

Sendo os migrantes positivamente selecionados e a desigualdade de renda um fator determinante na decisão de migrar, podemos nos questionar quais os resultados da saída desses indivíduos na desigualdade entre as regiões. Em [Santos \(2007\)](#) encontra que a migração gera um aumento das rendas médias de todos os estados (com exceção de São Paulo e Rio de Janeiro) e do Brasil, ou seja, há um aumento para as regiões mais pobres e queda para as regiões mais ricas, indicando convergência de renda. Esse resultado é contrastante com o de [Júnior et al. \(2005\)](#), onde foi sugerido que o fluxo desses trabalhadores mais produtivos pode estar agravando a as desigualdades regionais de renda no Brasil.

A partir da literatura nacional e internacional é possível perceber a importância da dinâmica migratória para os diferenciais de renda tanto entre regiões quanto entre trabalhadores. A migração gera mudanças nos dois aspectos, quando comparados migrantes e não migrantes. Mas a literatura também aponta para outro fator importante nesse contexto: a migração de retorno. Nos voltaremos agora para esse objeto de estudo de modo a comparar um grupo maior de agentes, incluindo agora aqueles que decidem permanecer no destino e aqueles que optaram por voltar a sua origem.

## 2.2 Migração de retorno e autosseleção

A teoria econômica vem buscando explicações para o subprocesso da migração de retorno. Se os agentes escolhem de forma racional seu local de residência, então qualquer resultado posterior deve ser fruto da decisão inicial de migrar ([BLAVASCHI, 2016](#)). Dessa forma, uma das primeiras explicações para a remigração vem da teoria neoclássica que associa o retorno aos erros de previsão cometidos pelo migrante.

A teoria neoclássica vê o retorno como fracasso, onde houve um erro de precisão

dos custos de migração ou o capital humano não foi recompensado como esperado. A nova teoria da migração, por sua vez, vê o retorno como uma estratégia previamente calculada, onde foram alcançadas as metas (CONSTANT; MASSEY, 2002; DUSTMANN; KIRCHKAMP, 2002; DUSTMANN, 2003; CASSARINO, 2004). As duas teorias são contrastantes não só por defenderem que o retorno representa fracasso ou sucesso, mas argumentarem que a decisão de migrar é permanente ou apenas temporária.

Se a migração for temporária, então a região de origem pode se beneficiar das habilidades e do capital adquiridos pelos migrantes durante a experiência migratória (DUSTMANN; KIRCHKAMP, 2002). Essa contribuição do retornado vai depender da natureza da seleção em que estão inseridos. Se os migrantes de retorno forem positivamente selecionados isso representará um ganho para a região de origem, já que esses trabalhadores serão motivados e produtivos, mas se há seleção negativa, o retornado representa um grupo desmotivado, que tende a diminuir o bem estar na origem.

Com as hipóteses de que a migração é de natureza temporária e dinâmica, a literatura tem buscado entender as consequências do retorno em relação à sua natureza de seleção. Borjas e Bratsberg (1996) identificam uma relação entre a seleção que determina migração inicial e a aquela que ocorre na migração de retorno. Dessa forma, se os migrantes foram inicialmente positivamente selecionados, então os retornados serão menos qualificados, valendo o contrário.

Coulon e Piracha (2005) utilizaram o modelo de seleção proposto por Lee (1978) e Lee (1983) e o método semiparamétrico proposto por DiNardo, Fortin e Lemieux (1995) para construir as densidades salariais contrafactuais afim de determinar o padrão de seleção dos retornados da Albânia. Os autores identificaram seleção negativa dos remigrados, apesar da migração representar um aumento dos rendimentos em relação aos não migrantes.

Identificando uma lacuna nos estudos sobre migração de retorno entre México e Estados Unidos, Biavaschi (2016) utilizou dados dos censos americanos e mexicanos de 2000 para estimar as densidades salariais dos migrantes em dois cenários: com e sem migração de retorno. Para isso, a autora propõe um procedimento semiparamétrico que complementa o estimador apresentado por DiNardo, Fortin e Lemieux (1995) e utilizado em Chiquiar e Hanson (2005) apenas para características observáveis. Dessa forma, busca-se determinar como os retornados se comparam aqueles que permanecerem em relação as suas características observáveis e não observáveis.

Os resultados de Biavaschi (2016) indicam que apesar dos retornados serem menos favorecidos em características observáveis no mercado de trabalho, suas habilidades não observáveis parecem compensar esse problema. Sem migração de retorno

os mexicanos estariam ganhando mais nos Estados Unidos. Em termos de variáveis não observáveis, esses indivíduos estariam ganhando aproximadamente 7,7% a mais na mediana, e 4,5% na média. Esse trabalho também destaca que as diferenças entre retornados e permanentes estão nas características não observáveis, isto é, em características observáveis os retornados estariam se saindo pior, mas há evidências de que estes possuem características como empreendedorismo, ambição, entre outras, que os tornam bem sucedidos no país de acolhimento, em relação aos permanentes.

No Brasil temos alguns estudos indicando a natureza da seletividade dos retornados. [Ramalho e Queiroz \(2011\)](#) utilizaram os dados da PNAD de 2007 para estudar os salários dos trabalhadores remigrados. Os resultados sugerem seleção positiva dos migrantes permanentes em atributos não observados em relação aos não migrantes. Em relação aos retornados os migrantes permanentes apresentam vantagem em termos de fatores não observados. Apesar disso, também em fatores não observados, os remigrados se saem melhor que os não migrantes. Resumidamente, os remigrados estariam ganhando mais se tivessem permanecido, no entanto eles ainda auferem vantagem econômica ao retornar ao estado de naturalidade. Resultados semelhantes foram encontrados nos estudos de [Gama e Machado \(2014\)](#). Os autores identificam seleção positiva em habilidades não observáveis para os remigrados e migrantes permanentes em relação aos não migrantes, mas não comparam relações entre permanentes e retornados.

Ao destacar o estado da Bahia como principal emissor de migrantes no Nordeste, cujo principal destino (cerca de 58%) é o estado de São Paulo e grande parte dos dos migrantes paulistas terem como destino a Bahia, [Assis, Costa e Mariano \(2012\)](#) propuseram uma análise do impacto do estoque migratório brasileiro, tanto de migrantes como de retornados, sobre a distribuição de renda dos estados da Bahia e de São Paulo. Os principais achados indicam que mesmo que os migrantes residentes na Bahia possuam atributos produtivos superiores aos dos paulistas, estes últimos recebem remunerações médias mais altas devido ao efeito regional positivo do mercado de trabalho paulista sobre os salários. Essa superioridade se mantém em relação aos retornados, dessa vez, devido à superioridade dos atributos dos migrantes paulistas. No mercado paulista, se mantém a vantagem do não migrante, havendo indícios de que os migrantes corroboram as desigualdades de renda no estado em virtude da queda da renda média pelos salários inferiores desses indivíduos.

Outras evidências dos diferenciais salariais e de capital humano entre paulistas e migrantes são sugeridas por [Neto e Magalhães \(2004\)](#). Diferenciais de renda entre migrantes e paulistas tendem a ser inicialmente favoráveis aos naturais. Mas essas diferenças podem ser diminuídas ou até mesmo superadas a partir de ajustamentos pelos diferenciais educacionais e de origem. Além disso, o autor ressalta que esses

diferenciais de rendas entre migrantes e paulistas não tendem a desaparecer e por conseguinte há concentração de renda nesse mercado.

A migração de retorno pode representar a saída para o problema da fuga de cérebros gerando um ganho de capital humano para região de origem, como destacam [Mayr e Peri \(2008\)](#). No entanto, as políticas certas de incentivo devem ser empregadas. Além disso, a migração permanente e a migração de retorno exercem papel crucial na determinação dos diferenciais de renda nas regiões e entre elas. Dessa forma, com o avanço das técnicas econométricas e disponibilidades de dados, é de suma importância a contínua dedicação a trabalhos com o objetivo de determinar a natureza de seleção e as consequências do processo migratório no Brasil.

## 3 Migração interestadual no Brasil: fatos observados

Os deslocamentos populacionais são amplamente investigados na literatura porque além de serem motivados por questões econômicas como a busca por melhores condições de vida e ganhos de capital humano, também são responsáveis por uma série de mudanças no espaço, nas relações e na economia. No Brasil, manteve-se predominante um fluxo migratório destinado ao Sudeste do País, principalmente para o estado de São Paulo, que aos poucos foi se redelineando com a contribuição da migração de retorno das regiões receptoras, direcionada aos polos tradicionalmente fornecedores. Portanto, o objetivo deste capítulo é apresentar um breve histórico da migração no Brasil, destacando o papel da migração de retorno. Também registra evidências sobre padrões na migração interna a partir dos conceitos de migrante retorno e migrante permanente.

### 3.1 Tendências Migratórias no Brasil

A partir dos anos de 1940, no Brasil, predominou um fluxo direcionado às grandes capitais como São Paulo e Rio de Janeiro, impulsionado pelos desequilíbrios entre as regiões e pelo crescimento urbano-industrial centralizado que esses estados ofertavam. As regiões de expansão da fronteira agrícola também se destacaram por sua capacidade de absorção, o estado do Paraná e as regiões Centro-Oeste e Norte. A origem desses migrantes também foi predominante, vinham principalmente do Nordeste e Minas Gerais, grandes ofertantes de mão-de-obra não absorvida por suas próprias economias. A seca que assolou o Nordeste nos anos 50 e a melhora no sistema de transporte impulsionaram a saída de pessoas e tornaram essa região a maior emissora de migrantes. Segundo [Netto Júnior et al. \(2003\)](#), nos anos 50, metade dos migrantes era oriunda na região Nordeste, dos quais 70% tinham como destino o Sudeste. Os autores também destacam que a partir dessa década houve uma expansão da migração no Brasil de 9,8% do total da população em 1950, para 11,7% em 1960 e 13,2% em 1970. Na década de 1960, os desequilíbrios regionais e a hegemonia de São Paulo enquanto centro dinâmico da economia nacional manteve sem modificações significativas a dinâmica migratória vigente nas décadas anteriores, com ligeiras mudanças de posição. São Paulo toma a liderança como principal destino, seguido pelo Paraná, Goiás e Rio de Janeiro. O Nordeste também manteve-se como principal polo de expulsão ([NETTO JÚNIOR et al., 2003](#); [BRITO, 2016](#)).

A década de 1970 marca o início das transformações das tendências migratórias dominantes. Segundo [Baeninger et al. \(1999\)](#), os efeitos da desconcentração relativa das atividades econômicas iniciadas nessa década só foram percebidas pelos deslocamentos de população nos anos de 1981 a 1991, indicando uma defasagem entre as mudanças econômicas e as mudanças demográficas. Além do esgotamento das fronteiras agrícolas, a desconcentração da indústria foi uma das mudanças observadas no país na passagem dessas duas décadas. São Paulo, que em 1970 concentrava 58% da indústria de transformação, baixou para 51,9% em 1985, essa expansão para o interior do estado e para outros estados como Minas Gerais, Paraná, Bahia, dentre outros, acabou por modificar os fluxos migratórios nessas direções. Para [Cunha \(2003\)](#), há de se considerar também a crise econômica dos anos 80 sobre esses movimentos populacionais. Destacando a hipótese de [Pacheco \(1993\)](#), o autor salienta que esse evento pode ter incentivado não só uma redução da migração em direção aos grandes centros, mas também a migração de retorno devido à redução do crescimento econômico.

O fenômeno da migração de retorno já se fazia perceber a partir dos anos 70 e 80, saindo de 11% para 24,5% do total da migração nacional, respectivamente nas duas décadas. Esse movimento representava a mitigação dos estados perdedores e elevação dos ganhadores. Dados apresentados por [Cunha \(2016\)](#) mostram que de 1986 a 1991 São Paulo apresentou um fluxo significativo de emigrantes cujo destino era a origem. Somente o fluxo de retorno São Paulo - Região Nordeste apresentou 47% do total da migração de retorno. A expansão dos espaços de migração viria se confirmar nos anos 90. Nessa década reduziram-se os fluxos de longa distância, mantendo como polos de absorção os estados de São Paulo, Rio de Janeiro, Goiás e Distrito Federal ([BAENINGER, 2016](#)).

Segundo [Oliveira, Ervatti e O'Neill \(2011\)](#), no quinquênio 1995/2000 mantiveram-se as principais correntes migratórias, com destaque para migração de retorno no contrafluxo. O estados do Nordeste apresentaram o maior número de retornados entre os estados brasileiros, com destaque para o Ceará e a Paraíba, cuja origem de seus retornados foi, principalmente, São Paulo e Rio de Janeiro. [Siqueira \(2006\)](#) ressalta que para o estado historicamente mais procurado pelos migrantes esse período representou não só de perda de atração comparado com outras décadas, mas também uma saída massiva de migrantes cujo destino era a origem. Essa autora salienta que após um período de extensiva oferta de emprego e concentração industrial, São Paulo passava por um momento de reestruturação das suas atividades produtivas o que pode ter afetado o emprego em termos de oferta e condições, com um aumento do emprego informal. Paralelamente, eram crescentes os problemas de congestionamento populacional, que acabam gerando uma série de questões, como sobrecarga nos serviços públicos, violência e escassez de moradia.

Em uma análise regional, [Siqueira \(2006\)](#) mostra que no fim do século XX (1995-2000) o Norte (62.685), Sudeste (488.587) e o Centro-Oeste (261.971) do país apresentaram saldos migratórios positivos, enquanto Sul (– 19.195) e Nordeste (–784.048) apresentaram saldos negativos. Como destaque no Sudeste além do estado de São Paulo, está o estado de Minas Gerais que além de reduzir o volume de emigrantes, aumentou o número de imigrantes, crescimento que pode ter relação com a migração de retorno. No período de 1999-2004 houve uma reversão de alguns saldos. O Sudeste passou a apresentar saldo negativo (–218.479), o Sul apresentou saldo positivo (33.013). Já no Nordeste houve um arrefecimento do saldo negativo, diminuindo a perda líquida para – 86.662. A participação da migração de retorno voltou a crescer nesse período nas regiões em que houve reversão ou queda do saldo negativo, dessa forma observa-se uma influência desse movimento na dinâmica migratória brasileira.

## 3.2 Caracterização da migração interestadual

Nesta seção são apresentados dados mais recentes sobre a migração interestadual no Brasil, considerando o cruzamento das questões de estado de residência na data de entrevista, estado de residência anterior e estado de naturalidade. Conforme dados do Censo Demográfico Brasileiro de 2010, entende-se o migrante permanente como o indivíduo que declarou residência em uma unidade federativa na data da entrevista, tendo declarado como residência anterior o estado de nascimento. O migrante retornado, por sua vez é o indivíduo declarou residir no seu estado de nascimento, mas que anteriormente residia em outro estado. Tais conceitos permitem estimar estoques de migrantes para o período de 2000 a 2010.

A Tabela 1 contém dados do quantitativo de migrantes interestaduais segundo a unidade federativa de residência em 2010 – migrantes não retornados – coluna (1) e migrantes remigrados de São Paulo – coluna (3) – e unidade federativa de residência anterior – migrantes retornados – coluna (2) –. Observa-se que o estado de São Paulo concentra não somente a maior parcela de migrantes permanentes (23%), mas também é o estado de origem da maior parcela de migrantes retornados, ou seja, 24,28% dos remigrados para seu estado de nascimento residiam anteriormente no estado. Destarte, mesmo atraindo muitos migrantes, São Paulo também é um grande emissor para outros estados. Goiás aparece em segundo lugar como principal destino dos migrantes brasileiros com 8,37% de não naturais, sendo residência anterior de apenas 3,66% de remigrados.



Tabela 1 – Distribuição dos migrantes interestaduais segundo estado de residência na data censitária e estado de residência anterior - 2010

Unidade Federativa	(1) Migrantes Permanentes (Residência em 2010)	(2) Migrantes Retornados (Residência anterior)	(3) Retornados de SP (Residência em 2010)	(4) Migran- tes/Habitantes da UF (%)
Rondônia (RO)	117.844 (1,61)	12.154 (1,43)	418 (0,20)	7,54
Acre (AC)	23.566 (0,32)	1.976 (0,23)	139 (0,07)	3,21
Amazonas (AM)	127.458 (1,74)	10.090 (1,18)	162 (0,08)	3,66
Roraima (RR)	53.511 (0,73)	2.749 (0,32)	15 (0,01)	11,88
Pará (PA)	322.553 (4,41)	34.775 (4,08)	1.418 (0,69)	4,25
Amapá (AP)	74.006 (1,01)	4.834 (0,57)	68 (0,03)	11,05
Tocantins (TO)	142.891 (1,95)	14.840 (1,74)	502 (0,24)	10,33
Maranhão (MA)	142.511 (1,95)	19.911 (2,34)	3.684 (1,78)	2,17
Piauí (PI)	92.141 (1,26)	11.786 (1,38)	5.864 (2,84)	2,95
Ceará (CE)	137.176 (1,87)	17.325 (2,03)	12. (5,94)	1,62
Rio Grande do Norte (RN)	90.603 (1,24)	8.827 (1,04)	4.959 (2,40)	2,86
Paraíba (PB)	113.011 (1,54)	14.808 (1,74)	8.190 (3,96)	3,00
Pernambuco (PE)	186.934 (2,55)	27.123 (3,18)	17.210 (8,32)	2,13
Alagoas (AL)	73.259 (1,00)	12.245 (1,44)	7.088 (3,43)	2,35
Sergipe (SE)	75.160 (1,03)	7.914 (0,93)	2.848 (1,38)	3,63
Bahia (BA)	304.219 (4,16)	49.448 (5,81)	36.197 (17,50)	2,17
Minas Gerais (MG)	492.874 (6,74)	61.498 (7,22)	42.122 (20,37)	2,52
Espírito Santo (ES)	209.894 (2,87)	21.207 (2,49)	1.970 (0,95)	5,97
Rio de Janeiro (RJ)	471.007 (6,44)	58.553 (6,88)	6.649 (3,22)	2,95
São Paulo (SP)	1.684.486 (23,02)	206.792 (24,28)	-	4,08
Paraná (PR)	368.095 (5,03)	55.499 (6,52)	35.417 (17,13)	3,52
Santa Catarina (SC)	454.487 (6,21)	44.925 (5,28)	5.259 (2,54)	7,27
Rio Grande do Sul (RS)	114.216 (1,56)	22.770 (2,67)	5.468 (2,64)	1,07
Mato Grosso do Sul (MS)	159.592 (2,18)	22.189 (2,61)	4.028 (1,95)	6,52
Mato Grosso (MT)	304.940 (4,17)	34.758 (4,08)	4.028 (0,75)	10,05
Goiás (GO)	612.654 (8,37)	31.174 (3,66)	3.288 (1,59)	10,20
Distrito Federal (DF)	368.097 (5,03)	41.420 (4,86)	-	14,32
População estimada	7.317.185 (100)	851.592 (100)	206.791 (100)	

Fonte: Elaboração própria com base nos dados do Censo Demográfico de 2010.

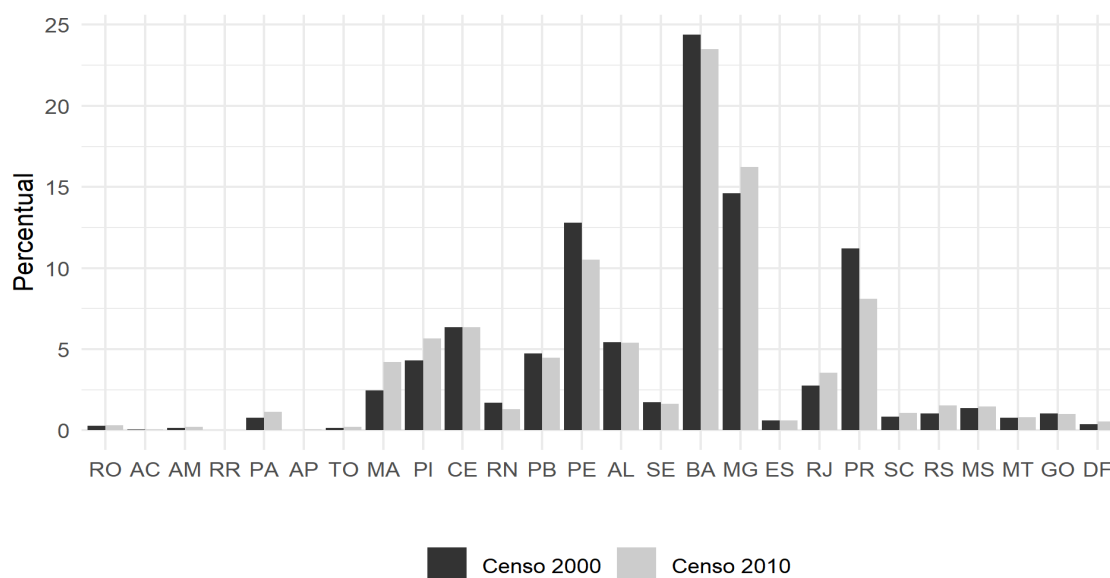
Nota: Valores expandidos para população. Percentual entre parênteses.



Dada a grande relevância do estado de São Paulo na atração de fluxos migratórios, em particular, como residência anterior de quase 1/4 dos remigrados no período analisado, a coluna (3) da tabela em destaque apresenta a distribuição dos migrantes retornados de São Paulo segundo seu estado de nascimento (residência em 2010). Destaca-se que do total de remigrados provenientes de São Paulo, cerca de 47% tiveram como destino os estados da região Nordeste e o estado de Minas Gerais, 20,37%.

Os dados apresentados na Figura 1 mostram entre que entre os Censos Demográficos de 2000 e 2010 não houve mudanças significativas nos padrões de representatividade dos estados de origem dos migrantes residentes em São Paulo (SP). Os cinco estados com maior população migrante em SP são os mesmos nos dois censos. Note-se que o estado da Bahia é o maior fornecedor nos dois períodos. O fluxo São Paulo–Bahia tem sido observado como recorrente nos estudos sobre mobilidade populacional. [Assis, Costa e Mariano \(2012\)](#), por exemplo, ressaltam que 60% dos migrantes retornados baianos residiam anteriormente em São Paulo, segundo dados da PNAD de 2009. Minas Gerais é o segundo estado mais representativo entre os migrantes, no censo de 2010 há um ligeiro aumento de envio. No entanto, no período de 2005 a 2010 apresentou equilíbrio na entrada e saída de migrantes, passando para categoria de estado com rotatividade migratória conforme [\(RAMALHO; FIGUEIREDO; NETTO JÚNIOR, 2016\)](#).

Figura 1 – Distribuição de migrantes em São Paulo por estado de origem - Censos de 2000 e 2010



Fonte: Elaboração própria a partir dos dados do Censo Demográfico de 2000 e 2010.  
Nota: Valores expandidos para população

São Paulo tem se sobressaído como absorvedor de migrantes há décadas e as tendências de origem dos migrantes também não sofreram significativas mudanças

ao longo do tempo. O que vem se destacando é a grande perda de população sofrida pelo estado, fruto especialmente da migração de retorno de nordestinos e de mineiros. O fluxo de migrantes oriundos de São Paulo para a região Nordeste constitui-se basicamente de migrantes retornando aos seus estados, o que vinha contribuindo para saldos migratórios positivos. Os dados apresentados acima sugerem que esse movimento continua na mesma direção.

Em se tratando de capital humano, a Tabela 2 relata os percentuais de migrantes interestaduais residentes no estado de São Paulo em 2010 e de acordo com a região de nascimento (residência anterior) e o nível de escolaridade. Observa-se uma maior concentração de migrantes com baixa instrução de origem na região Nordeste, 69,50% dos migrantes sem instrução são oriundos dessa região. Esse padrão se repete para os níveis fundamentais e médio com 67,14% e 56,27%, respectivamente. Nos níveis de instrução mais altos encontram-se em destaque os sulistas com 43% dos migrantes com ensino superior e aqueles oriundos da região Centro-Oeste do país (23,63%). Dos migrantes que possuem pós-graduação, 43,97% são oriundos da região Sul. No geral, a população migrante em São Paulo é de baixa instrução, com cerca 53% de pessoas no nível mais baixo, apenas 6,04% dessa população declarou possuir nível superior e 0,73% disse possuir pós-graduação.

Tabela 2 – Distribuição dos migrantes permanentes em São Paulo segundo faixa de instrução e por região de residência anterior (nascimento)

Nível de instrução	Região de origem do migrante					Total
	Norte	Nordeste	Sudeste	Sul	Centro Oeste	
Sem-instrução	29.345	616.956	13.706	151.099	76.555	893.472
%	3,31	69,50	1,54	17,02	8,62	53,04
Nível Fundamental	10.058	202.769	6.309	53.271	29.594	303.580
%	3,33	67,14	2,09	17,64	9,80	18,02
Nível Médio	15.976	204.166	9.159	85.817	45.792	363.873
%	4,43	56,57	2,54	23,78	12,69	21,06
Nível superior	7.530	21.012	3.755	42.656	23.753	101.739
%	7,63	21,29	3,80	43,22	24,06	6,04
Pós-Graduação	1.017	2.131	700	5.221	2.806	12.298
%	8,56	17,95	5,89	43,97	23,63	0,73
Indefinido	324	6.099	232	2.028	812	9.524
%	3,41	6,23	2,44	21,36	8,55	0,56

Fonte: Elaboração própria com base nos dados do Censo Demográfico 2010.

Nota: Valores expandidos para população

As informações trazidas nesse capítulo são úteis para identificar as tendências

e padrão de migração interestadual no Brasil. Chama atenção o estado de São Paulo enquanto grande polo de absorção e de residência anterior de remigrados. No entanto, a influência da migração de retorno assumiu papel relevante nos movimentos populacionais do Brasil. Uma questão relevante é investigar os efeitos da migração de retorno sobre a desigualdade salarial do grande quantitativo de migrantes residentes em São Paulo. As seções a seguir procuram responder como esse problema pode ser investigado.

## 4 Metodologia

Considerando que o objetivo central deste trabalho é investigar o impacto da migração interestadual de retorno sobre a distribuição de salários dos migrantes que permaneceram no estado de São Paulo, precisamos recuperar a densidade salarial de todos os migrantes interestaduais (inclusive daqueles que estiveram em São Paulo), tendo em vista que os salários são observados apenas para aqueles que permaneceram no estado.

### 4.1 Modelo Empírico

Seguindo [Biavaschi \(2016\)](#), a decisão de ficar na região de destino ou retornar para a região de nascimento vai depender do valor do benefício líquido da permanência. Caso ele seja maior do que zero, o trabalhador deve permanecer na região de destino, caso contrário deve remigrar. Dessa forma, fazemos  $S_i = 1$  um indicador que o trabalhador migrante permaneceu em São Paulo e  $S_i = 0$  caso tenha retornado para o estado de nascimento. A decisão em destaque será determinada por:

$$S = \begin{cases} 1 & Z_i' \alpha > \epsilon_i \\ 0 & Z_i' \alpha \leq \epsilon_i \end{cases} \quad \text{para } i = 1, \dots, r + n, \quad (4.1)$$

Onde  $r$  é o número de retornados;  $n$  de migrantes permanentes;  $Z_i$  uma matriz de variáveis que influenciam na decisão de ficar no destino;  $\alpha$  um vetor de parâmetros (inclusive intercepto) e  $\epsilon_i$  um termo randômico. Portanto,  $(Z_i' \alpha - \epsilon_i)$  mensura o benefício líquido da decisão de permanência.

A equação de determinação de salários do tipo *minceriana* para um migrante selecionado aleatoriamente na população de migrantes presentes em São Paulo é dada por:

$$Y_i^* = X_i' \beta + u_i^* \quad i = 1, \dots, r + n, \quad (4.2)$$

Onde  $Y_i^*$  é o logaritmo do salário-hora dos migrantes;  $X_i$  é o conjunto de variáveis observáveis que determinam o processo de pagamentos de salários;  $\beta$  um vetor de parâmetros (inclusive intercepto) e  $u_i^*$  é o termo aleatório que mensura atributos produtivos não observáveis na determinação desses rendimentos.

Alguns pressupostos devem ser admitidos em relação às equações 4.1 e 4.2. Primeiro,  $(Y_i^*, S_i, X_i, Z_i)$  são variáveis aleatórias observadas. Segundo, devemos assu-

mir que  $(X_i, Z_i, u_i^*, \epsilon_i)$  são variáveis aleatórias independentemente e identicamente distribuídas (i.i.d) e  $(X_i, Z_i)$  são variáveis aleatórias exógenas. Terceiro, os termos randômicos  $u_i^*$  e  $\epsilon_i$  podem ser correlacionados, pois características não observáveis podem influenciar tanto a decisão de permanência do migrante quanto o seu salário.

Cabe ressaltar que o salário  $Y_i$  é observado apenas para os migrantes permanentes. Logo:

$$Y_i = S_i Y_i^* \quad i = 1, \dots, r + n \quad (4.3)$$

Desejamos obter a distribuição de salários  $Y_i^*$  para todos os migrantes, incluindo salários dos remigrados caso tivessem permanecido na região de destino, isto é,  $f(Y_i^*)$ . No entanto, tal distribuição é dada por um deslocamento da função de densidade  $f(u_i^*)$  pelo valor médio de salários preditos na região de destino  $X_i' \beta$  (BIAVASCHI, 2016). Portanto, precisamos de métodos para estimativas não tendenciosas para a função de densidade de probabilidade  $f(u_i^*)$  e dos vetores de parâmetros  $\beta$ .

Usamos duas estratégias de identificação conforme discutido a seguir. No primeiro caso, usamos o método de identificação no infinito (CHAMBERLAIN, 1986; HECKMAN, 1990) para estimar  $f(u_i^*)$ , uma vez que não observamos  $u_i^*$  para os trabalhadores remigrados. Na segunda etapa, dado o problema de viés de seleção amostral, estimamos o modelo 4.1 e 4.2 com o uso de restrição de exclusão, isto é, uma variável instrumental que não determina diretamente a formação de salários, mas se relaciona com a decisão de permanência na região de destino.

#### 4.1.1 Estimação dos parâmetros

O problema de seleção amostral acontece quando as observações disponíveis não provêm de uma amostra aleatória da população. Há evidências de que o grupo de migrantes brasileiros são positivamente selecionados em relação aos não migrantes (JÚNIOR et al., 2005; FREGUGLIA; PROCÓPIO, 2013; GAMA; MACHADO, 2014). Da mesma forma, temos que os migrantes de retorno se autosselecionam na decisão de retornar e isto deve ser considerado para que estimativas não tendenciosas do resultado no mercado de trabalho do migrante interestadual sejam produzidas (CHISWICK, 1999; BORJAS; BRATSBERG, 1996; BIAVASCHI, 2016).

O modelo bivariado proposto por Heckman (1979), conhecido como *Tobit-2*, é um dos modelos mais utilizados em problemas de seleção amostral. Ele pode ser representado pela equação de seleção 4.1 e a equação de resultado potencial 4.2. Sua estimativa por Máxima Verossimilhança pressupõe os termos aleatórios  $u_i^*$  e  $\epsilon_i$  seguem uma distribuição normal bivariada. No entanto, como essa suposição está sujeita a erros de especificação de distribuição, modelos que relaxem essa hipótese têm sido sugeridos

na literatura especializada (CAMERON; TRIVEDI, 2005; TOOMET; HENNINGSEN et al., 2008; WOJTYS; MARRA; RADICE, 2016).

O uso de funções cópulas, por exemplo, assegura flexibilidade distributiva ao relaxar a hipótese de normalidade conjunta de  $u_i^*$  e  $\epsilon_i$ , e garante eficiência na estimação das equações 4.1 e 4.2. Além disso, a opção oferecida pela abordagem de cópula é útil sempre que a precisão das estimativas dos parâmetros estruturais for a prioridade.

Como visto na equação 4.3, a variável  $S_i$  controla se a variável de resultado  $Y_i^*$  é observada ou não. Seja  $F_i$  a função de distribuição cumulativa (FDC) conjunta de  $(S_i, Y_i^*)$  e  $F_{1i}$  e  $F_{2i}$  a FDC marginal de  $(S_i, Y_i^*)$ , respectivamente. Assumimos a normalidade das distribuições marginais, a dependência entre elas, por sua vez, é modelada usando a abordagem de cópula. Logo,  $S_i \sim \mathcal{N}(\mu_{1i}, 1)$  (modelo probit) e  $Y_i^* \sim \mathcal{N}(\mu_{2i}, 1)$ , onde  $\mu_{1i}$  e  $\mu_{2i}$  são preditores lineares e  $\sigma > 0$ .  $F_{1i}$  e  $F_{2i}$  estão relacionadas à equação de seleção e de salários respectivamente. Assim, a distribuição conjunta é determinada da seguinte forma:

$$F_i(S_i, Y_i^*) = C(F_{1i}(S_i), F_{2i}(Y_i^*); \theta), \quad (4.4)$$

Onde  $\theta$  é um parâmetro de dependência e  $C$  é uma função cópula bidimensional. As famílias de cópulas implementadas neste estudo são as arquimedianas: *Clayton*, *Frank*, *Ali-Mikhail-Haq (AMH)*, *Farlie-Gumbel-Morgenstern (FGM)*, *Gumbel* e *Joe*, bem como as versões rotacionadas ( $90^\circ$ ,  $180^\circ$  e  $270^\circ$ ) para *Clayton*, *Joe* e *Gumbel*. A cópula elíptica Gaussiana juntamente com covariadas lineares ou pré-especificadas não lineares corresponde ao modelo de Heckman (1979) usando o método de máxima verossimilhança (WOJTYS; MARRA; RADICE, 2016). Mais detalhes sobre as características das famílias selecionadas encontram-se na Tabela A2 do apêndice, que contém as equações e os intervalos do parâmetro de dependência para cada família de cópula.

Para as cópulas normais, *Frank*, *FGM* e *AMH* o teste para viés de seleção de amostra pode ser baseado no parâmetro de dependência  $\theta$ , pois a ausência de viés de seleção de amostra é equivalente à condição  $\theta = 0$  e  $\theta = 1$  para a cópula de *Gumbel*. As cópulas *Clayton*, *Joe* e *Gumbel* não têm uma interpretação direta de  $\theta$ . Nesses casos a estatística de *Kendall*, que apresenta uma relação matemática com  $\theta$ , pode facilitar a interpretação e comparação do grau de dependência obtido por diferentes cópulas<sup>1</sup>. Esse parâmetro geralmente situa-se no intervalo  $[-1, 1]$ , e indica independência entre variáveis aleatórias modeladas na cópula. Quanto mais próximo  $\tau$  estiver de -1, mais forte será a associação negativa, valores mais próximos de 1 indicam forte dependência positiva (BRECHMANN; SCHEPSMEIER, 2013; MARRA; WYSZYNSKI, 2016).

Outra vantagem das cópulas é que elas permitem uma especificação de modelo por partes, isto é, as distribuições marginais não são limitadas a pertencer à mesma

<sup>1</sup> Por exemplo, para cópula normal o parâmetro  $\tau$  de *Kendall* em termos de  $\theta$  é obtido através da expressão  $\frac{2}{\pi} \arcsin(\theta)$ , mais detalhes da relação entre os parâmetros  $\tau$  e  $\theta$  podem ser consultados em Marra e Wyszynski (2016), página 6.

família da distribuição de cópula bivariada escolhida. Utilizar uma versão rotacionada, disponível para *Clayton*, *Joe* e *Gumbel*, é útil quando a cobertura da cópula não é completa. A dependência negativa entre as equações só pode ser capturada pelas versões rotacionadas a 90° e 270° (MARRA; WYSZYNSKI, 2016). Quando rotacionadas a 180° obtêm - se as cópulas de sobrevivência (BRECHMANN; SCHEPSMEIER, 2013).

Com a função cópula incorporada à estimação dos parâmetros das equações 4.1 e 4.2 temos que a equação log-verossimilhança a ser maximizada será a seguinte (WOJTYS; MARRA; RADICE, 2016):

$$\ell = \sum_{i=1}^n \left\{ (1 - S_i) \log F_{1i}(0) + S_i \log \left( f_{2i}(Y_i^*) - \frac{\partial}{\partial Y} F_i(0, Y^*) \Big|_{Y^* \rightarrow Y_i^*} \right) \right\}. \quad (4.5)$$

Essa função é uma soma de dois conjuntos disjuntos da amostra: um para aqueles valores não observados do salários, ou seja, que representam os migrantes retornados, e outro para as demais observações. Utilizando 4.4, temos:

$$\ell = \sum_{i=1}^n \{ (1 - S_i) \log F_{1i}(0) + S_i \log (f_{2i}(Y_i^*) (1 - z_i)) \}, \quad (4.6)$$

Onde  $z_i = \frac{\partial}{\partial v} C(F_{1i}(0), v; \theta) \Big|_{v \rightarrow F_{2i}(Y_i^*)}$  e  $F_{1i}$  e  $F_{2i}$  são funções de distribuição cumulativas marginais<sup>2</sup> de  $(S_i, Y_i^*)$ , respectivamente.

Em muitos estudos variáveis contínuas como idade e anos de estudo podem guardar uma relação não linear com as variáveis de interesse, já que envolvem questões como produtividade e ciclo de vida. Nesse casos, a relação é feita de forma pré-especificada o que não garante que a complexidade dessa relação seja devidamente capturada e entendida. Dessa forma, o modelo bivariado de seleção amostra com cópula pode flexibilizar a modelagem de covariadas contínuas utilizando funções suaves que são representadas por *splines*<sup>3</sup> de regressão (MARRA; RADICE, 2010; MARRA;

<sup>2</sup> A normalidade dessas funções implica em  $F_{1i}(0) = \Phi(-\mu_{1i})$  e  $f_{2i}(Y_i^*) = \sigma^{-1} \phi((Y_i^* - \mu_{2i}) \sigma^{-1})$

<sup>3</sup> Funções suaves são utilizadas com o intuito de flexibilizar o relacionamento entre as variáveis de interesse e variáveis explicativas contínuas e assim evitar as desvantagens da abordagem paramétrica. Essas funções suaves são representadas por *splines* (funções polinomiais particionadas). O *spline* de regressão de uma determinada variável explicativa é composto por um combinação linear de funções básicas conhecidas  $b_{jk}(x_j^*)$  e parâmetros de regressão desconhecidos  $\delta_{jk}$ . Logo,  $s_j(x_j^*) = \sum_{k=1}^{q_j} \delta_{jk} b_{jk}(x_j^*)$ , onde  $j$  indica o termo suave da  $j$ -ésima variável explicativa e  $q_j$  é o numero de parâmetros de regressão (bases do *spline*) usados para representar o  $j$ -ésimo termo suave, o cálculo dessa expressão para cada  $j$  produzirá  $q_j$  curvas que multiplicadas pelo vetor de parâmetros e somadas produzirão estimativas lineares ou não lineares para  $s_j(x_j^*)$ . A base  $q$  escolhida para uma função suave determina o grau de flexibilidade máxima para um termo suave e bases devem possuir propriedades matemáticas e numéricas adequadas (MARRA; RADICE, 2010; WOJTYS; MARRA; RADICE, 2016).



WYSZYNSKI, 2016; WOJTYS; MARRA; RADICE, 2016). Portanto, as matrizes dos modelos de seleção e salários podem ser construídas como a soma de componentes paramétricos (intercepto, *dummies* e variáveis categóricas) e funções suaves desconhecidas das covariáveis contínuas  $K_{1,2}$ . Logo, as equações 4.1 e 4.2, na abordagem semiparamétrica, podem ser reescritas da seguinte forma:

$$S = \begin{cases} 1 & Z'_i \alpha + \sum_{k_1=1}^{K_1} s_{1k_1}(z_{1k_1i}) > \epsilon_i \\ 0 & Z'_i \alpha + \sum_{k_1=1}^{K_1} s_{1k_1}(z_{1k_1i}) \leq \epsilon_i \end{cases} \quad \text{for } i = 1, \dots, r+n, \quad (4.7)$$

$$Y_i^* = X'_i \beta + \sum_{k_2=1}^{K_2} s_{2k_2}(z_{2k_2i}) + u_i^* \quad i = 1, \dots, r+n, \quad (4.8)$$

Onde  $s_{1k_1}$  e  $s_{2k_2}$  são os termos suavizados desconhecidos dos regressores contínuos das equações de seleção e salário, respectivamente. Tanto a abordagem totalmente paramétrica quando esta abordagem são discutidas neste estudo.

A definição da cópula utilizada é feita através de testes que consistem em estimar vários modelos e selecionar aquele com melhor critério de informação. Utiliza-se o critério de penalização por parcimônia *Bayesiano Akaike* ou *Schwarz* (AIC e BIC, respectivamente), o que permite identificar a cópula e o modelo que melhor se ajusta aos dados (WOJTYS; MARRA; RADICE, 2016)<sup>4</sup>.

#### 4.1.2 Estimação de densidade contrafactual

A estratégia de recuperação da densidade contrafactual de salários  $f(Y_i^*)$ , isto é, da densidade salarial dos migrantes caso não houvesse migração de retorno, foi dividida em duas etapas. Na primeira foi recuperada a distribuição das características não observáveis, isto é,  $f(u_i^*)$ . Nesse ponto apenas as características produtivas não mensuráveis são levadas em conta. Note que não é possível recuperar essa distribuição diretamente da equação de salários 4.2, tendo em vista que os resíduos  $u_i^*$  apenas poderiam ser calculados para os trabalhadores que decidiram permanecer em São Paulo. Logo, não podemos determinar, a partir dos dados, o componente não observável da equação de salários ao incluir remigrados na amostra (“devolvê-los para São Paulo”). Para superar essa dificuldade utilizamos a técnica proposta por Biavaschi (2016).

Considerando que a distribuição  $f(u_i^*)$  pode ser escrita a partir do pressuposto da Lei da Probabilidade Total como a soma ponderada da distribuição dos termos

<sup>4</sup> Os critérios AIC e BIC podem ser calculados, respectivamente, por  $AIC = -2\hat{\ell} + 2k$  e  $BIC = -2\hat{\ell} + \log(n)k$ , onde  $\hat{\ell}$  é o valor maximizado da função de log-verossimilhança e  $k$  o grau de liberdade do modelos de seleção e salário. Ambos os critérios penalizam a adição de variáveis nas estimações (WOJTYS; MARRA; RADICE, 2016)



de erro nas subamostras de permanentes e retornados cujos pesos são dados pela probabilidade de estar em qualquer subamostra, temos que:

$$f(u_i^*|Z_i'\alpha) = f(u_i^*|S_i = 1, Z_i'\alpha) \Pr(S_i = 1|Z_i'\alpha) + f(u_i^*|S_i = 0, Z_i'\alpha) \Pr(S_i = 0|Z_i'\alpha) \quad (4.9)$$

Como explicado, o componente não observável ( $f(u_i^*|S_i = 0, Z_i'\alpha)$ ) dessa soma é desconhecido. No entanto, se a probabilidade de ficar em São Paulo  $\Pr(S_i = 1|Z_i'\alpha)$ , condicionada às variáveis que influenciam essa decisão, for próxima de 1, então:  $f(u_i^*|Z_i'\alpha) = f(u_i^*) \approx f(u_i^*|S_i = 1, Z_i'\alpha) = f(u_i^*|S_i = 1)^5$ , então o viés desaparece e a distribuição contrafactual de fatores não observáveis pode ser estimada diretamente nessa subamostra na qual quase todos os indivíduos são permanentes. A intuição por trás dessa estratégia é que a seleção desaparece no limite para aqueles indivíduos no conjunto de alta probabilidade (CHAMBERLAIN, 1986). Devemos adicionar dois pressupostos básicos para que essa técnica garanta uma estimativa consistente da quantidade de interesse, são eles: o conjunto de características que seleciona a amostra ( $Z$ ) deve conter variáveis exógenas e as observações precisam ser i.i.d.

Essa estratégia é conhecida como identificação no infinito e tem sido utilizada na literatura para estimar o termo constante em modelos de seleção semiparamétrica<sup>6</sup>. O trabalho de Biavaschi (2016) foi o primeiro a aplicar essa estratégia para recuperação de distribuição contrafactual com objetivo de considerar a seleção em características não observáveis em uma densidade contrafactual. O presente estudo buscou aplicar essa estratégia na estimativa da distribuição das características não observáveis baseando-se em uma subamostra selecionada nas características observadas ( $Z$ ), na qual quase todos os indivíduos permanecem no estado de São Paulo.

Seja  $H_i$  o indicador se a observação faz parte do conjunto de alta probabilidade e seja  $H_i = 1 [\Pr(S_i = 1|Z_i'\hat{\alpha}) > \bar{p}_n]$ , o estimador proposto para  $f(u_i^*)$  será:

$$\widehat{f(u_i^*)} = \frac{\sum_{i=1}^n \frac{1}{h} K\left(\frac{u - u_i^*}{h}\right) S_i H_i}{\sum_{i=1}^n S_i H_i}, \quad (4.10)$$

Onde  $K(\cdot)$  é um estimador de densidade de *Kernel* da variável aleatória  $u_i^*$  para um conjunto de observações cuja probabilidade de estar na amostra selecionada é próxima de 1 e  $h$  é o parâmetro *bandwidth* ou largura dos intervalos de classe. Seguindo DiNardo, Fortin e Lemieux (1995), Chiquiar e Hanson (2005) e Biavaschi (2016) será utilizado o *Kernel* gaussiano com parâmetro de suavização ótimo dado por  $h = 1,06\hat{\sigma}N^{-1/5}$

<sup>5</sup> Assumindo  $f(u_i^*|S_i = 1, Z_i'\alpha) = f(u_i^*|S_i = 1)$  e  $f(u_i^*|Z_i'\alpha) = f(u_i^*)$

<sup>6</sup> Em trabalhos como o de Heckman (1990), Andrews e Schafgans (1998), Liu et al. (2009), Chzhen e Mumford (2011) e Shen (2013).

(SILVERMAN, 2018), onde  $\hat{\sigma}$  é o desvio-padrão da distribuição gaussiana,  $N$  o total de observações e  $\bar{p}_n$  é o valor que define o percentil que limita os indivíduos no grupo de alta probabilidade. Seguindo Biavaschi (2016), esse limite será o percentil 95º calculado a partir da predição de probabilidade pela equação de seleção (4.1).

Cabe observar que para determinar  $H_i$  precisamos antes da estimativa consistente de  $\hat{\alpha}$  para determinar as observações com alta probabilidade de permanência por meio da predição  $\Pr(S_i = 1|Z_i'\hat{\alpha})$  com a equação 4.2.

A segunda etapa da estimação contrafactual da densidade de salários dos migrantes consistiu em recuperar a distribuição completa dos salários somando a  $f(u_i^*)$  pela média do valor predito ( $X_i'\hat{\beta}$ ). Em outras palavras, os salários previstos reais e contrafactuais que foram calculados como o produto do retorno sobre as habilidades incluídas na equação de salários e as características dos migrantes permanentes (população total migrante) para distribuição real (contrafactual) prevista dos salário, isto é,  $\hat{\beta}X_j$ , onde  $j$  = apenas permanentes (população total migrante – devolvendo os remigrados para São Paulo). Nesse ponto são levadas em conta tanto características não observáveis quanto as observáveis na análise das diferenças entre retornados e permanentes. Essa separação é útil para identificar em que tipo de características se concentram as diferenças entre os migrantes.

## 4.2 Base de Dados e Tratamento

Neste estudo foram utilizados os microdados do Censo Demográfico 2010. Essa base de dados permite identificar as diferentes categorias de migrantes a partir do cruzamento dos dados sobre lugar de residência, lugar de nascimento, lugar de última residência e duração de residência.

A amostra selecionada para a análise empírica é composta de homens com idade entre 18 e 70 anos que estavam empregados no momento do recenseamento. Em relação à posição do domicílio foram incluídos na amostra os chefes de domicílio e cônjuges<sup>7</sup> que não estivessem frequentando algum curso, essa estratégia foi aplicada afim de excluir migrantes agregados. Também foram excluídos da análise os indivíduos que não declaram cor e que não responderam o questionário referente à migração, além dos funcionários públicos, militares e sem remuneração, essa exclusão foi feita de modo que a análise esteja voltada para decisões de migração e mercado de trabalho (GAMA; MACHADO, 2014; LIMA; SIMÕES; HERMETO, 2015).

Após os cortes mencionados a amostra final totalizou 28.458 migrantes interestaduais, sendo deste total 24.421 (85,81%) migrantes que ficaram no estado de São Paulo e 4.037 (14,18%) migrantes retornados.

Define-se migrante de retorno aquele que morava, na data do recenseamento, em seu estado de naturalidade, tendo declarado residência anterior no estado de São Paulo. O migrante não retornado é definido como aquele que, na data da entrevista, declarou residir fora de seu estado de naturalidade (residência em São Paulo). Existem dois critérios para definir a condição migratória: última etapa e data fixa (curto prazo). No critério de data fixa o indivíduo é questionado sobre o local de residência há cinco anos da entrevista. Com o objetivo de captar um fluxo de retornados em maior espaço de tempo foi utilizado o critério última etapa<sup>8</sup> combinado ao quesito lugar de nascimento para identificação dos migrantes.

Este estudo busca aplicar as análises para diferentes grupos educacionais, os indicadores educacionais disponíveis nos dados do censo referentes ao nível de instrução são: sem instrução e fundamental incompleto, fundamental completo e médio incompleto, médio completo e superior incompleto e superior completo. O modelo de seleção foi construído a partir de quatro grupos de variáveis. O primeiro grupo refere-se às características pessoais do indivíduo como idade, raça, posição e nível de instrução. O segundo grupo diz respeito aos vínculos familiares como: posição no

<sup>7</sup> A partir dos dados do Censo Demográfico de 2010 não é possível identificar o parentesco direto de um indivíduo cuja posição no domicílio seja filho ou enteado. Dessa forma a amostra é limitada aos responsáveis e cônjuges de modo que se torne viável construir a variável de restrição de exclusão (filho nascido em SP).

<sup>8</sup> A amostra identifica a duração de residência de permanentes e retornados de até 9 anos.

domicílio (chefe), chefe de família casado, e chefe casado com cônjuge nascido em SP, chefe com filho menor de 14 anos e número de pessoas no domicílio, como essas variáveis estão relacionadas aos custos de migração e vínculos à pessoas no destino elas devem funcionar bem como preditores da decisão de permanecer. Outro grupo de covariadas utilizado é o relacionado a emprego e ocupação do indivíduo, como posição no trabalho e se é beneficiário de programa social. Por fim, variáveis referentes à residência como setor do domicílio e região de origem do migrante<sup>9</sup>.

A variável filho nascido no destino, isto é, no estado de São Paulo, foi incluída como restrição de exclusão, e deve ser incluída na equação de seleção e excluída da equação de salários. É esperado que ter um filho nascido em São Paulo seja um forte preditor da decisão de permanência, já que se trata de uma *proxy* para vínculo social no destino, o que aumenta o custo de oportunidade do retorno. Por outro lado, não é provável que o local de nascimento do filho de um indivíduo seja determinante no processo salarial (BIAVASCHI, 2016).

O modelo de salários também foi construído a partir dessas variáveis educacionais, socioeconômicas e de estrutura familiar (exceto pela restrição de exclusão), incluindo a variável possuir esposa no destino, pois essa variável deve capturar os efeitos das redes de sociais (BIAVASCHI, 2016). A Tabela A1 do apêndice contém um resumo das variáveis disponíveis na base de dados relacionadas à migração, às condições socioeconômicas e educacionais que foram utilizadas utilizadas nas regressões. Vale ressaltar que esse conjunto de variáveis está de acordo com a literatura nacional e internacional especializada conforme a disponibilidade do Censo Demográfico de 2010 (MINCER, 1978; SANTOS, 2007; RAMALHO; QUEIROZ, 2011; JUSTO; FERREIRA, 2012; GAMA; MACHADO, 2014; BIAVASCHI, 2016).

A Tabela 3 contém as estatísticas descritivas das características socioeconômicas e demográficas amostra de trabalhadores selecionados de migrantes permanentes, retornados e dos não migrantes naturais do estado de São Paulo. Há importantes diferenças entre esses grupos que merecem destaque. Primeiro, os migrantes permanentes são, em média, mais jovens e menos escolarizados que os migrantes retornados, fato que pode indicar uma seleção positiva dos migrantes de retorno. Em relação aos não migrantes, tanto migrantes retornados quanto permanentes são, em média, mais jovens e menos escolarizados. Há também uma forte concentração de migrantes na faixa educacional sem instrução, esse fato e a maior escolaridade dos paulistas em relação aos migrantes é consistente com os achados de Neto e Magalhães (2004) para o período de 1991-2000.

<sup>9</sup> Uma limitação dos dados do Censo Demográfico consiste na impossibilidade de identificação do momento (antes ou após o retorno) em que as característica dos indivíduos como, por exemplo, a obtenção de benefício social, foram adquiridas.

Tabela 3 – Médias das características socioeconômicas e demográficas observáveis de migrantes e não migrantes (Amostra)

Variável	Permanentes	Migrantes Retornados	Não Migrantes
Não branco (%)	59,94	52,14***	31,71
Idade	32,97	39,33***	42,39
Sem instrução (%)	55,46	49,27***	36,64
Nível fundamental (%)	20,09	19,64	18,51
Nível médio (%)	21,43	24,85***	32,68
Nível superior (%)	2,88	5,97***	11,67
Pós graduação (%)	0,14	0,27	0,49
Chefe (%)	79,64	81,60***	81,32
Chefe e vive com cônjuge (%)	61,30	68,62***	71,23
Filho de 14 anos (%)	62,67	70,25***	72,22
Filho de 14 anos nascido em SP (%)	37,72	28,78***	71,52
Chefe e vive cônjuge nascido em SP (%)	14,58	7,38***	79,42
Emprego com carteira assinada (%)	75,34	45,65***	61,05
Emprego sem carteira assinada (%)	11,42	19,97***	10,87
Autônomo (%)	12,59	30,64***	24,45
Programa social (%)	0,58	1,09***	0,25
Empregador (%)	0,65	3,74***	3,63
Nº de pessoas no domicílio	3,47	3,36***	3,45
Renda no trabalho principal (R\$)	1074,26	1109,75**	1596,45
Salário-hora (R\$)	6,31	6,50**	9,36
Setor urbano (%)	93,31	87,71***	91,09
Setor rural (%)	6,69	12,29***	8,91
Origem - Sudeste (%)	19,43	25,29***	-
Origem - Nordeste (%)	65,44	40,72***	-
Origem - Centro-Oeste (%)	2,52	3,86***	-
Origem - Norte (%)	1,10	0,82*	-
Origem - Sul (%)	11,51	29,30***	-
Amostra total	24.421	4.037	140.166

Fonte: Elaboração própria com base nos dados do censo 2010.

Nota: Níveis de significância: \*10%, \*\* 5%, \*\*\* 1% para um teste t para diferenças nas médias entre os grupos de permanentes e retornados. A amostra consiste em homens de 18 a 70 anos, empregados em atividade remuneratória. Amostra não expandida, isto é, não foi considerado o peso amostral.

Nota-se que em relação à raça, entre os migrantes predominam os não brancos, enquanto que para os paulistas apenas 31,71% declarou raça diferença da branca. Em relação à família, cerca de 61% dos migrantes permanentes são chefes de família casados, e cerca de 62% tem filho com idade até 14 anos no domicílio, para os retornados essas proporções aumentam para 68,62% e 70,25%. Mesmo constituindo um custo para migração, essas variáveis também se relacionam de forma não observada com a decisão de permanência, de forma que os custos psicológicos ou não mensuráveis devem ser levados em conta (SJAASTAD, 1962; NIKOLKA, 2018).

Em relação ao emprego e ocupação, percebe-se uma predominância de trabalhadores com carteira de trabalho assinada no grupo de permanentes, enquanto para o grupo de retornados há uma distribuição maior entre empregados com carteira

assinada, empregadores, autônomos e trabalhadores sem carteira de trabalho assinada. A hipótese de que o migrante retorna após a adquirir recursos para empreender no pós-retorno pode justificar essa distribuição mais heterogênea entre os migrantes retornados (JUSTO; FERREIRA, 2012). Além disso, segundo Dustmann e Kirchkamp (2002), as atividades empreendedoras dos migrantes retornados são um benefício da migração temporária, tendo em vista que podem contribuir para geração de riqueza e emprego.

Em relação à renda no trabalho principal verifica-se que entre migrantes e paulistas há uma diferença salarial considerável. Essa diferença pode ser explicada em termos das características observáveis, tendo em vista que os atributos pessoais paulistas tendem a favorecê-los no mercado de trabalho. Para Neto e Magalhães (2004), esses diferenciais estão associados ao capital humano do migrante que é influenciado pelas características do estado de origem.

As variáveis demográficas indicam que para permanentes e retornados há grande predominância da origem no Nordeste, principalmente para o grupo de permanentes. Diversos trabalhos já destacaram não só a preferência dos nordestinos pelo destino São Paulo, mas também que há um crescente movimento de retorno para essa região (SACHSIDA et al., 2009; SANTOS, 2007; RAMALHO; QUEIROZ, 2011; OLIVEIRA; ERVATTI; O'NEILL, 2011; RAMALHO; FIGUEIREDO; NETTO JÚNIOR, 2016, 2016).

## 5 Resultados

Neste capítulo são apresentados os principais resultados referentes aos determinantes da decisão de remigração e salários (seção 5.1), resultados das estimativas de densidade real e contrafactual (seção 5.3) e os resultados para análise em diferentes grupos educacionais (5.4). Na seção 5.5 é verificada a consistência dos achados a partir de modelos alternativos e teste de validade instrumental.

### 5.1 Determinantes da decisão de remigrar e salários

Uma série de especificações empíricas para o modelo de determinação conjunta da decisão de remigração e salários (equações 4.1 e 4.2) foram previamente estimados considerando vários tipos de cópulas, covariadas e formas paramétricas e semiparamétricas. A Tabela A3 do Apêndice contém os valores do critério de informação AIC. Os resultados sugerem que a cópula *Joe*<sup>1</sup> com rotação 180° se ajusta melhor aos dados em todas as especificações. O modelo mais completo, contendo variáveis referentes às características pessoais, familiares, trabalho e residência obteve o menor valor AIC<sup>2</sup>. A especificação semiparamétrica não favorece uma cópula diferente dos outros modelos, mas apresenta menor valor AIC em relação à especificação totalmente paramétrica. O parâmetro de dependência das distribuições marginais de 4.1 e 4.2 registra valor  $\theta = 2,068$ , estatisticamente diferente de zero, cujo parâmetro  $\tau$  de *Kendall* é 0,37, sugerindo que fatores não observados que afetam a decisão de permanência também afetam os salários positivamente.

Os resultados da Tabela 4 referem-se ao modelo de seleção amostral bivariado semiparamétrico baseado na cópula *Joe* 180°. Em se tratando da restrição de exclusão: trabalhador que reside com filho nascido em São Paulo. A expectativa teórica é que tal variável tenha uma relação importante na predição da decisão de permanência do indivíduo, pois se trata de um forte vínculo familiar. Segundo os resultados do modelo de seleção essa variável é significativa a 1% e indica que ter um filho nascido em São Paulo pode aumentar a probabilidade de o trabalhador migrante permanecer no estado. Ter cônjuge nascido em São Paulo também deve aumentar essa probabilidade, evidenciando o papel dos laços familiares formados no destino.

<sup>1</sup> Especificações totalmente paramétricas, utilizando a cópula *Joe* 180° e a cópula Normal (equivalente ao modelo de Heckman (1979) em dois estágios estimado por máxima verossimilhança, Tobit-2), são disponibilizadas no apêndice. A partir da figuras A1 verifica-se que os resultados são muito próximos aos da abordagem semiparamétrica, mantendo-se as conclusões do modelo selecionado.

<sup>2</sup> A partir de simulações, Wojtys, Marra e Radice (2016) mostram que a escolha da cópula verdadeira é melhor executada a partir o critério AIC.

Tabela 4 – Estimativas das equações de seleção e salários

Covariadas	Equação de Seleção ( <i>probit</i> , $S = 1$ ) (1)	Equação de Salários (2)
<i>Não tem filho (categoria omitida)</i>		
Filho de 14 anos nascido em SP	0,3392*** (0,0256)	-
Filho de 14 anos	-0,4561*** (0,0397)	-0,0001 (0,0093)
<i>Branco (categoria omitida)</i>		
Não Branco	-0,0287 (0,0215)	-0,0408*** (0,0063)
<i>Sem instrução (categoria omitida)</i>		
Nível Fundamental	-0,1990*** (0,0277)	0,1006*** (0,0079)
Nível Médio	-0,3194*** (0,0270)	0,2098*** (0,0080)
Nível superior	-0,3072*** (0,0518)	0,7989*** (0,0187)
Pós-Graduação	-0,2986 (0,2184)	1,1171*** (0,0788)
<i>Cônjuge (categoria omitida)</i>		
Chefe	0,2716*** (0,0462)	0,0692*** (0,0110)
<i>Chefe solteiro (categoria omitida)</i>		
Chefe e vive com cônjuge	-0,4472*** (0,0426)	-0,0189* (0,0099)
Chefe e vive cônjuge nascido em SP	0,5253*** (0,0372)	0,0161* (0,0091)
<i>Emprego sem carteira assinada (categoria omitida)</i>		
Emprego com carteira assinada	0,5858*** (0,0285)	0,0761*** (0,0095)
Autônomo	-0,0472 (0,0329)	0,1749*** (0,0122)
Empregador	-0,4008*** (0,0791)	0,4421*** (0,0382)
<i>Não recebe Programa Social (categoria omitida)</i>		
Recebe Programa Social	-0,4308*** (0,1089)	-0,1046*** (0,0393)
<i>Setor rural (categoria omitida)</i>		
Setor Urbano	0,2314*** (0,0349)	0,1956*** (0,0122)
<i>Origem - Sudeste (categoria omitida)</i>		
Origem - Norte	0,1929* (0,1066)	0,0141 (0,0292)
Origem - Nordeste	0,2925*** (0,0262)	-0,0802*** (0,0080)
Origem - Sul	-0,3495*** (0,0307)	0,0076 (0,0112)
Origem - Centro-Oeste	-0,1337** (0,0587)	0,0353* (0,0200)
Intercepto	0,8442*** (0,0572)	1,3668*** (0,0182)
Idade ( <i>spline</i> )	sim	sim
Total de pessoas no domicílio ( <i>spline</i> )	sim	sim
Observações	28.458	24.421
Parâmetro de Dependência (Intervalo de confiança 95%): $\theta = 2.068(1.932, 2.209)$		
$\tau$ de Kendall : $\tau = 0.37$		

Fonte: Elaboração própria com base nos dados do censo 2010.

Nota 1: Erros-Padrão entre parêntese. : O modelo bivariado é do tipo cópula *Joe* rotacionada a 180°. Níveis de significância: \*10%, \*\* 5%, \*\*\* 1%



Chefes de família são mais propensos a permanecer em relação aos que ocupam a posição de cônjuge no domicílio. Por outro lado, ser chefe de família vivendo com cônjuge e ter filho com idade até 14 anos tende a diminuir a probabilidade de ser migrante permanente. Apesar da relação dessas variáveis com custos de migração (MINCER, 1978), esse resultado corrobora as evidências de que a presença de crianças em idade escolar aumenta as chances de retorno (NIKOLKA, 2018). Para a decisão de remigrar, a variável referente à raça não é significativa.

As variáveis referentes ao nível de instrução indicam uma relação inversa com a decisão do trabalhador não remigrar. Comparados àqueles sem instrução (categoria omitida), indivíduos com nível médio e superior tem menos chances de permanecer. A grande predominância de indivíduos sem instrução na amostra (55,46%) pode indicar que esse resultado está relacionado à frustração das expectativas de salários de migrantes com alto nível de instrução. Ao se deparar com um mercado de trabalho acirrado e seletivo esse grupo pode sofrer frustração em relação às suas expectativas de pagamento do capital humano, facilitando o retorno de trabalhadores mais instruídos. Já os trabalhadores sem instrução podem acessar uma maior oferta de empregos de baixa qualidade. Essa suposição é consistente com Ambrosini et al. (2010), onde defende-se que as decisões de migração e retorno estão relacionadas ao prêmio salarial oferecido para uma classe específica de habilidades na origem e no destino.

As variáveis referentes à condição de ocupação no trabalho indicam que ser autônomo não tem efeito sobre a probabilidade de permanência, enquanto a condição de empregador aumenta as chances do indivíduo ser um migrante retornado. Em contrapartida, manter um emprego com carteira de trabalho assinada pode aumentar as chances de permanecer em São Paulo. Justo e Ferreira (2012) destacam que alguns indivíduos buscam adquirir recursos no local de destino com o intuito de empreender posteriormente no local de nascimento. A variável programa social, que identifica o indivíduo que obteve na rendimento de programas como Bolsa Família ou PETI (Programa de Erradicação do Trabalho Infantil), indica uma relação inversa à decisão de permanência, isto é, receber transferência de programas do governo pode aumentar a probabilidade de retorno do indivíduo que migrou para São Paulo.

Migrantes com origem no Norte e Nordeste apresentam maior propensão de permanecer em São Paulo. Migrantes nascidos na região Sul ou no Centro-Oeste, por sua vez, têm mais chances de retornar. Apesar de um movimento mais intenso de retorno para o Nordeste, a migração ainda é fortemente relacionada aos níveis de pobreza e desigualdade de renda nas regiões tradicionalmente emissoras, sendo esse um fator de expulsão populacional (GAMA; MACHADO, 2014). Paralelamente, a dinâmica migratória brasileira sofreu significativas mudanças, a região Centro-Oeste passou a ser uma região de grande absorção e retenção de migrantes, e São Paulo passou a

apresentar perdas migratórias para essa região e para o Sul do país (BAENINGER, 2016).

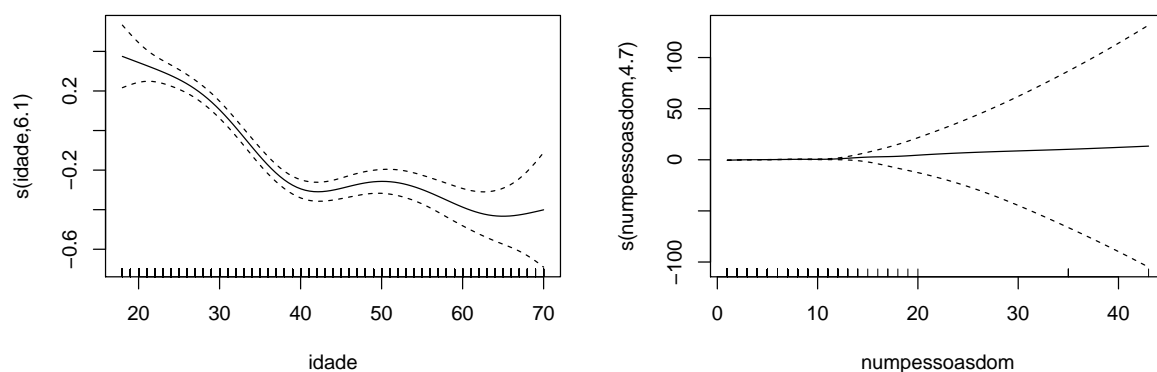
Outros achados foram obtidos na estimação da equação de salário. O logaritmo do salário-hora responde positivamente ao nível de instrução. Quanto maior esse nível, maiores os coeficientes da equação de salários. Note-se ainda que a variável referente à raça aponta para discriminação no mercado de trabalho, com uma diferença desfavorável aos não brancos de 4%. Em termos de vínculos familiares temos que, um chefe de família cujo cônjuge é natural de SP auferir salários maiores em relação a um chefe de família solteiro. Esse fato pode estar relacionado às redes sociais construídas no destino (BLAVASCHI, 2016). A diferença de salários entre trabalhadores que recebem auxílio de programa social e trabalhadores que não recebem é de -9,93%. Alguns estudos defendem que essas políticas podem gerar incentivos adversos como o desestímulo ao trabalho (TEIXEIRA, 2011; BRITO, 2011). Cabe também ressaltar que os trabalhadores ocupados como empregados com carteira assinada, autônomos e empregadores auferem salários mais altos em relação aos trabalhadores empregados sem carteira de trabalho assinada (categoria omitida).

Residir em área urbana aumenta os rendimentos em 21,96%. Com relação à origem do migrante foram significativas as variáveis de nascimento no Nordeste e Centro-Oeste. Migrantes com origem no Nordeste (Centro-Oeste) auferem salários inferiores (superiores) em relação aos de origem Sudeste. Esse resultado pode ser explicado em termos de atributos educacionais tendo em vista a concentração de migrantes nordestinos com níveis de escolaridade mais baixos. Como visto na Tabela 2, 69,5% dos migrantes sem instrução são oriundos dos estados do Nordeste, enquanto o Centro-Oeste se destaca nos níveis mais elevados de instrução.

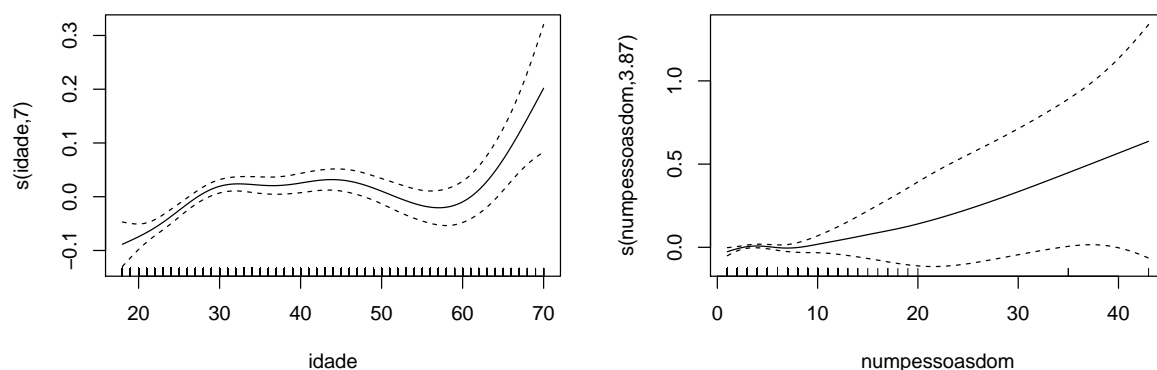
A Figura 2 registra os resultados não paramétricos das equações de seleção e salários para o modelo cópula selecionado. Os resultados para variável idade são consistentes com a suposição de tanto rendimentos quando a probabilidade de permanência respondem não linearmente à medida que as pessoas envelhecem. A estimativa suavizada da equação seleção indica que a probabilidade de permanecer no destino tende a diminuir à proporção que a idade aumenta, até por volta dos 40 anos, quando há uma ligeira mudança para uma tendência positiva até por volta dos 50 anos, voltando a diminuir. O resultado não paramétrico para os salários sugere que o salário aumenta com a idade até por volta dos 30 anos, se tornando quase constante dos 30 aos 40 anos e declinando após essa faixa até os 60 anos. A partir desse ponto a relação é crescente. Os intervalos do resultado suavizado para o número de pessoas no domicílio (`numpessoasdom`) na equação de seleção e salários contêm linha zero para a maior parte da faixa de valores da covariável, indicando que esse pode ser um preditor fraco da permanência e dos rendimentos.

Figura 2 – Resultados das estimativas não paramétricas das equações de seleção e salários

(a) Equação de Seleção



(b) Equação de Salários



Fonte: Elaboração própria a partir dos dados do Censo Demográfico de 2010

Nota: Estimativas das funções suaves do modelo bivariado cópula Joe 180°. Intervalos de confiança a 95%. O "tapete" na parte inferior dos gráficos representa os valores das covariadas. Os números mostrados no eixo y e em cada gráfico indicam os graus estimados de liberdade das curvas suaves. P-valores para os termos idade e número de pessoas no domicílio da equação de seleção (a): 0,000, 0,000, respectivamente; equação de salários (b): 0,000, 0,035.

## 5.2 Perfil do Grupo de Controle

Este estudo usa a estratégia de identificação ao infinito para recuperar a distribuição de salários contrafactual em características não observadas, isto é, a distribuição que vigoraria caso não houvesse migração de retorno (HECKMAN, 1990; ANDREWS; SCHAFGANS, 1998; BIAVASCHI, 2016). Conforme discutido nas equações 4.9 e 4.10, essa distribuição pode ser estimada livre de viés de seleção amostral usando um grupo de controle formado por migrantes com Alta Probabilidade de Permanência (GAP) na

região de destino. A Tabela 5 contém as médias das características observáveis para os indivíduos do GAP, isto é, aqueles trabalhadores migrantes cuja probabilidade predita de permanência em São Paulo se encontra acima do percentil 95º (vide equação 4.1 e Tabela 4). Os dados permitem identificar as características que compõem o perfil desses trabalhadores.

Tabela 5 – Características médias dos migrantes do GAP versus demais migrantes

Variável	GAP ( $H_i = 1$ )	Outros migrantes ( $H_i = 0$ )
Não branco (%)	64,49	58,54***
Idade	24,09	34,39***
Sem instrução (%)	70,32	53,75***
Nível médio (%)	9,07	22,59***
Nível fundamental (%)	20,60	19,99
Nível superior (%)	0,00	3,49***
Pós graduação (%)	0,00	0,17***
Chefe (%)	92,90	79,24***
Chefe e vive com cônjuge (%)	43,95	63,30***
Chefe e vive cônjuge nascido em SP (%)	40,37	12,15***
Filho de 14 anos (%)	38,40	65,08***
Filho de 14 anos nascido em SP (%)	37,76	36,38
Emprego com carteira assinada (%)	99,16	69,65***
Autônomo (%)	0,28	15,93***
Empregador (%)	0,00	1,15***
Recebe Programa Social (%)	0,00	0,68***
Total de pessoas no domicílio	4,43	3,41***
Urbano (%)	98,31	92,21***
Origem - Sudeste (%)	3,16	21,16***
Origem - Nordeste (%)	95,71	60,15***
Origem - Centro-oeste (%)	0,14	2,85***
Origem - Norte (%)	0,98	1,07
Origem - Sul (%)	0,00	14,77***
Amostra total	1.422	27.036

Fonte: Elaboração própria com base no dados do Censo 2010.

Nota 1 :  $H_i$  indica se a observação encontra-se acima do percentil 95º.

Nota 2: Níveis de significância: \*10%, \*\* 5%, \*\*\* 1% para um teste de diferenças nas médias entre os grupos.

O primeiro ponto importante a ser destacado é a grande concentração de indivíduos no GAP cuja origem é a região Nordeste. Ou seja, 95% daqueles migrantes com alta probabilidade de permanência são oriundos dessa região. Esses indivíduos também apresentam nível de instrução baixo, com maior proporção de indivíduos na categoria sem instrução, 70,32%, nenhum migrante com nível superior ou pós graduação foi identificado, nessa amostra, como pertencente ao conjunto. Eles ainda são mais jovens em relação àqueles migrantes fora do grupo de controle, com uma

diferença de idade de 10,3 anos. Os vínculos familiares também se destacam nas diferenças entre esses grupos. Os indivíduos no GAP são, em média, mais propensos a ter um cônjuge nascido em São Paulo, em relação aos outros, assim como um ter filho nascido no destino. Em contrapartida, são, na maioria, chefes de família solteiros e sem filhos.

Em relação ao emprego e ocupação, o GAP é composto quase inteiramente por indivíduos com carteira assinada, a despeito do fato de que são indivíduos com baixíssima instrução. Esse fato será trazido novamente quando forem analisadas as diferenças entre os migrantes permanentes e de retorno em relação às suas características no cenário contrafactual. O grupo de indivíduos fora do conjunto contém, em média, mais indivíduos na posição de autônomos e empregadores, fato que pode estar relacionado às características não observáveis como empreendedorismo. Por fim, é importante destacar que, em média, o número de pessoas no domicílio é maior para aqueles indivíduos com alta propensão de serem migrantes permanentes, essa característica está relacionada aos custos de migração.

Conforme colocado na seção 4.1.2, calculamos os resíduos da equação de salários (vide Tabela 4) para grupo de controle (GAP) e usamos o estimador de núcleo *Kernel* (vide 4.10) para estimar a distribuição contrafactual de salários em características não observadas. Em seguida, conforme Biavaschi (2016), modificamos tal densidade pela média da distribuição de salários previstos para toda população migrante, o que corresponde à estimativa completa da distribuição dos salários dos migrantes interestaduais de São Paulo caso não houvesse migração de retorno. A seção seguinte relata as densidades dessas distribuições em ambos os cenários, além de comparar com a distribuição dos não migrantes paulistas.

### 5.3 Efeitos da migração de retorno sobre os salários dos imigrantes

A Tabela 6 resume o comportamento das densidades de salários por horas de trabalho estimadas para os migrantes conforme as métricas de decis, médias e diferenças de percentis (valores em logaritmo). A coluna (1) – fatual – foi obtida a partir da amostra de migrantes permanentes (residentes em São Paulo) com a aplicação do estimador de densidade *Kernel*. Sua primeira parte corresponde à densidade de salários-hora preditos em características observadas  $f(X'\hat{\beta})$ . A segunda resulta da estimativa de densidade sobre a distribuição dos resíduos  $f(Y_i - X'\hat{\beta})$ , que corresponde aos fatores não observáveis. Na terceira parte, tem-se a estimativa de densidade de salários totais, isto é, considerando a distribuição dos resíduos e o deslocamento pelos salários médios previstos. A coluna (2) – contrafactual – foi calculada no cenário em que não há migração

de retorno. Assim, no primeiro painel foi utilizada a amostra completa de migrantes interestaduais caso todos tivessem permanecido (salários previstos devolvendo os retornados para São Paulo) para recuperação da distribuição em variáveis observadas. No segundo painel, os valores foram obtidos a partir do cálculo de resíduos de salários para grupo de alta probabilidade de permanência (grupo de controle), como visto na seção 4. Já o terceiro painel refere-se aos salários totais, considerando a distribuição dos resíduos e o deslocamento pelos salários médios previstos. Em outras palavras, essa distribuição corresponde à soma das características observáveis médias ( $\hat{Y}_i$ ) com o componente não observado  $\hat{u}_i$  da população migrante em cada decil.

Em termos de características observáveis os migrantes interestaduais em São Paulo estariam, em média, ganhando mais caso não houvesse migração de retorno. Note-se que, na Tabela 6 a diferença logarítmica ao longo dos decis é quase sempre positiva, exceto pelos primeiros decis daqueles indivíduos na parte inferior da distribuição, para esses indivíduos há evidências de seleção negativa em observáveis. No cenário contrafactual, considerando apenas diferenças observáveis, a população migrante estaria recebendo 0,96% a mais, em média, e 0,41% na mediana.

O último painel da Tabela 6 mostra os decis da distribuição total de salários-hora. É possível ver que há uma influência mais evidente das características não observadas no efeito da migração de retorno para os decis extremos. Quando reintroduzidos no mercado de trabalho paulista, os retornados acabam aumentando o salário médio da população migrante em quase toda distribuição. Quando não consideramos características inatas, somente a parte média e superior da distribuição mostram melhora. As características não observáveis acabam melhorando o desempenho da cauda inferior e piorando parte da cauda superior, sugerindo que aqueles com rendimentos mais baixos se diferem dos demais em termos de características como: ambição, empreendedorismo, agressividade, propensão ao risco, motivação e etc. Destarte, a parcela de migrantes na parte inferior da distribuição estaria ganhando mais caso não houvesse migração de retorno. Ademais, a população total migrante estaria ganhando 4,28% a mais na média e 5,9% na mediana, devido à diferenças não observáveis entre migrantes permanentes e retornados.

As diferenças entre os grupos de migrantes permanentes e migrantes retornados mudam com a introdução de uma análise baseada em características não mensuráveis. Apesar do grupo de trabalhadores de baixa qualificação ser menos favorecido no mercado de trabalho, suas habilidades não mensuráveis parecem compensar essa falha. O que não ocorre para parte do grupo com habilidades observáveis superiores.

Tabela 6 – Decis das distribuições de salários dos migrantes com e sem migração de retorno

	(1) Fatural	(2) Contrafatural	Diferença (2)-(1)
Observáveis	$\hat{Y}_i$		
1º decil	1,1536	0,7506	-0,4031
2º decil	1,5105	1,5008	-0,0097
3º decil	1,5695	1,5657	-0,0039
4º decil	1,6127	1,6131	0,0003
5º decil	1,6562	1,6602	0,0041
6º decil	1,7102	1,7185	0,0083
7º decil	1,7694	1,7846	0,0152
8º decil	1,8583	1,8877	0,0294
9º decil	3,0303	4,1146	1,0843
Média	1,6915	1,7011	0,0096
Não observáveis	$\hat{u}_i$		
1º decil	-1,7295	-0,8846	0,8449
2º decil	-0,4898	-0,4099	0,0798
3º decil	-0,3129	-0,2257	0,0872
4º decil	-0,1751	-0,0962	0,0789
5º decil	-0,0554	-0,0059	0,0494
6º decil	0,0747	0,0956	0,0209
7º decil	0,2364	0,2148	-0,0216
8º decil	0,5056	0,4205	-0,0851
9º decil	1,9032	1,9032	0,0000
Média	0,0000	0,0332	0,0332
Total	$Y_i$		
1º decil	-0,0381	0,8164	0,8545
2º decil	1,2017	1,2911	0,0894
3º decil	1,3786	1,4754	0,0968
4º decil	1,5164	1,6048	0,0885
5º decil	1,6361	1,6951	0,0590
6º decil	1,7662	1,7966	0,0304
7º decil	1,9279	1,9158	-0,0120
8º decil	2,1971	2,1216	-0,0755
9º decil	3,5946	3,6042	0,0096
Média	1,6915	1,7343	0,0428
Diferenciais Salariais			
10º percentil - 90º percentil	1,1276	0,9466	-0,1810
10º percentil - 50º percentil	0,4776	0,4544	-0,0232
50º percentil - 90º percentil	0,6499	0,4923	-0,1576

Fonte: Elaboração própria a partir dos microdados do Censo Demográfico de 2010.

Nota 2: As predições de salários foram baseadas nos parâmetros estimados para o modelo bivariado do tipo cópula Joe rotacionada a 180°.

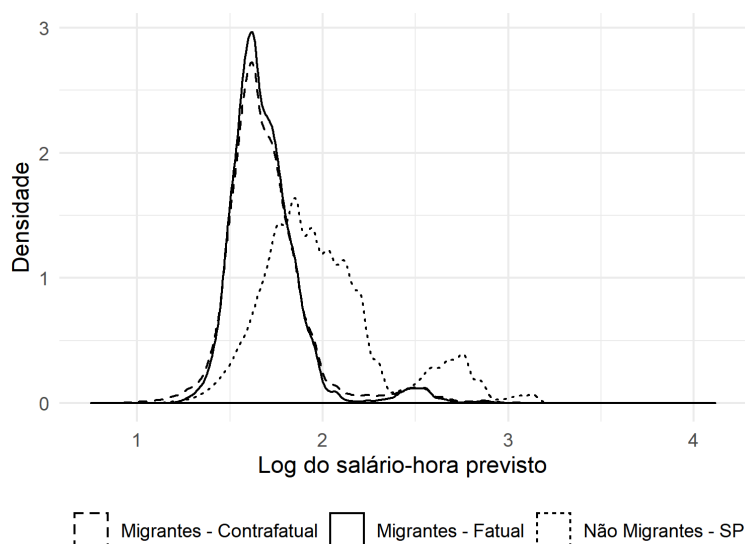
As Figuras 3 e 4 descrevem os resultados das distribuições de salários reais (linha sólida) e contrafatuais (linha pontilhada) graficamente. As diferenças estatisticamente não nulas entre essas linhas representam os efeitos da migração de retorno sobre a densidade dos salários dos imigrantes (DINARDO; FORTIN; LEMIEUX, 1995;



COULON; PIRACHA, 2005).

A Figura 3 mostra as distribuições condicionadas ao conjunto de variáveis observáveis. É possível observar que essas distribuições são muito próximas, indicando que retornados e permanentes são bastante semelhantes em termos das suas habilidades mensuráveis. A distribuição contrafactual da população total migrante mostra mais massa na parte inferior e menos massa na parte superior da distribuição quando comparada a distribuição real dos residentes em SP. A figura também revela que há uma significativa lacuna entre os rendimentos de migrantes e naturais em ambos os cenários, esse resultado é consistente com as conclusões de Neto e Magalhães (2004) e Batista e Cacciamali (2009).

Figura 3 – Densidade real e contrafactual do salário-hora para migrantes e paulistas - observáveis



Fonte: Elaboração própria a partir dos dados do Censo Demográfico de 2010.

Nota 1: A distribuição real representa a distribuição dos salários dos trabalhadores migrantes em São Paulo, considerando apenas as características observáveis. A distribuição contrafactual representa a distribuição dos salários caso todos os migrantes tivessem permanecido em SP, isto é, da população total migrante.

Nota 2: Um teste K-S rejeitou a hipótese nula de que as distribuições sejam iguais, a 1% de significância: estatística D = 0.0283 e p-valor = 0,0000.

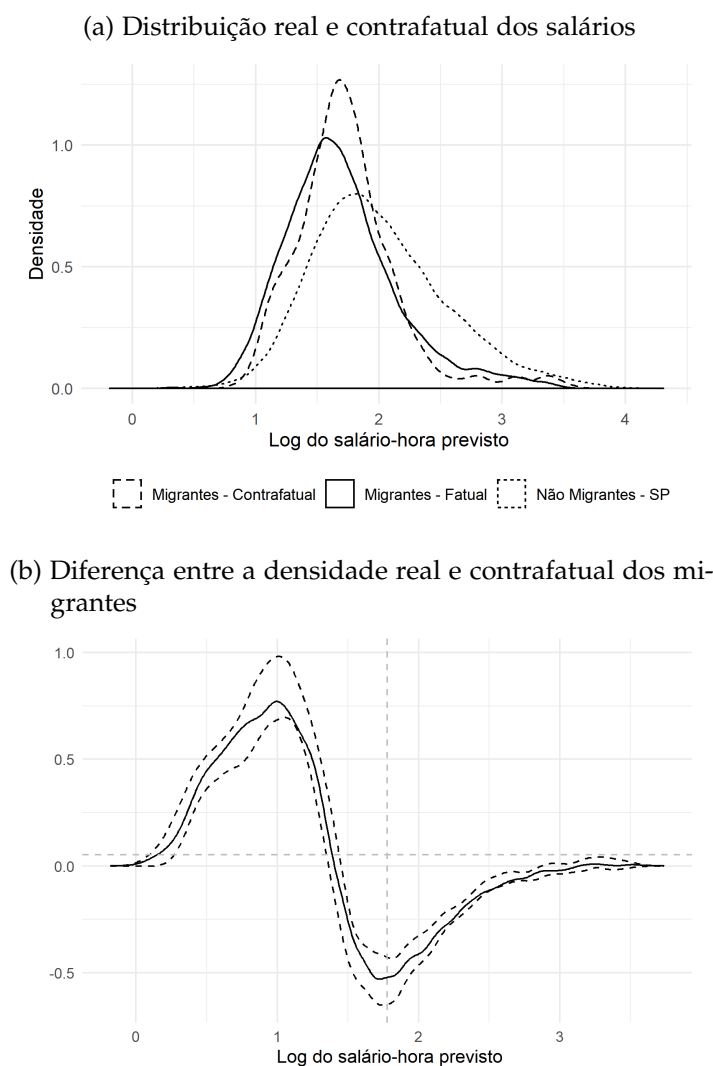
Nota 3: As predições de salários foram baseadas nos parâmetros estimados para o modelo bivariado do tipo cópula *Joe* rotacionada a 180°.

Segundo Batista e Cacciamali (2009) há um distanciamento entre migrantes e não migrantes na região sudeste. Neto e Magalhães (2004), por sua vez, apontam para magnitude dos diferenciais de renda dentro do universo de migrantes e não migrantes no estado de São Paulo devido aos diferenciais de capital humano. Já os autores Assis, Costa e Mariano (2012), ao analisar a migração entre os estados de São Paulo e Bahia identificam, mais uma vez, um hiato salarial entre baianos e paulistas, favorável ao último grupo.



A Figuras 4 (a) e (b), por sua vez, se referem as estimativas de distribuição de salários totais. Elas evidenciam maiores diferenças entre as distribuições. Na ausência de migração de retorno mais migrantes apareceriam na metade superior da distribuição, além de haver um deslocamento à direita na parte esquerda da distribuição, aumentando o salário médio dessa população. Como constatado na análise dos decis, parte da cauda superior dessa distribuição possui menos massa no cenário em que não há migração de retorno, indicando uma piora salarial. Essa figura dá maior suporte à seleção positiva em não observáveis dos migrantes de retorno oriundos de São Paulo.

Figura 4 – Densidade real e contrafactual do salário-hora total para migrantes e paulistas



Fonte: Elaboração própria a partir dos dados do Censo Demográfico de 2010

Nota 1 : (a) A distribuição real corresponde à distribuição dos salários totais (características observáveis e não observáveis) dos migrantes permanentes (residentes em SP), o contrafactual corresponde a distribuição dos salários totais dos migrantes interestaduais caso não houvesse migração de retorno, i.e., caso todos perdessem no destino.

Nota 2: (b) Intervalo de confiança representado pelas linhas pontilhadas, supondo um nível de 95% de confiança. A estimação dos erros-padrão foi feita usando a técnica de reamostragem *bootstrap*.

Nota 3: O teste *Kolmogorov - Smirnov* rejeitou-se a hipótese nula de que as densidades reais e contrafatuais sejam iguais, ao nível de 1% de significância: estatística  $D = 0.0955$  e  $p\text{-valor} = 0,0000$

Nota 4: As predições de salários foram baseadas nos parâmetros estimados para o modelo bivariado do tipo cópula *Joe* rotacionada a  $180^\circ$ .

Há evidências de que uma análise baseada apenas em características observáveis subestima o efeito da migração de retorno sobre a distribuição dos salários da população migrante. Além disso, observa-se que ao contar com não mensuráveis a distância entre as distribuições de migrantes de paulistas diminui em relação à análise em características observáveis. As diferenças entre as distribuições reais e contrafatuais são melhor visualizadas a partir da Figura 4 (b). Observa-se que as diferenças entre as densidades são maiores nos níveis mais baixos e médios de rendimento, indicando um efeito maior da migração de retorno.

A Tabela 6 mostra que a migração de retorno também afeta a desigualdade salarial entre os migrantes que permanecem no estado, a partir da diferença entre os percentis 10-90, 10-50 e 50-90 das distribuições reais e contrafatuais. A diferença salarial entre os percentis extremos diminuiria caso não houvesse migração de retorno, representando uma variação de  $-18,1\%$  nos diferenciais de salários. Na parte superior da distribuição também há um movimento de arrefecimento das desigualdades, caso não houvesse migração de retorno a diferença entre os percentis 90 e 50 seria  $-15,76\%$ . Na parte inferior da distribuição (10-50) há pouco efeito da migração de retorno sobre as diferenças salariais.

No geral, a desigualdade salarial entre a população migrante diminuiria se todos os migrantes decidissem ficar em São Paulo. Tendo em vista que sem migração de retorno, a média salarial dos migrantes aumenta na parte inferior e diminui na parte superior. Portanto, há uma tendência de aprofundamento da desigualdade entre os migrantes permanentes devido à saída de trabalhadores positivamente selecionados, aumentando a dispersão dos salários, considerando paralelamente a grande proporção de trabalhadores sem instrução que tendem a permanecer no destino.

A partir dos resultados das estimativas das distribuições dos salários reais e contrafatuais podemos fazer uma análise comparativa com a distribuição dos não migrantes nascidos em São Paulo. Na Tabela 7 estão dispostas as diferenças das distribuições salariais entre naturais e migrantes por decil em cada cenário. Como, geralmente, migrantes e não migrantes tendem a se diferenciar em termos de características não observadas, os diferenciais estão divididos por tipo de característica.

Em termos de características observáveis, a migração de retorno melhora a desigualdade salarial na parte inferior da distribuição. Considerando as características não observáveis (salário total), temos que a migração de retorno piora a desigualdade entre migrantes e não migrantes em quase todos os decis. No geral, há um aumento da desigualdade salarial em decorrência da saída de trabalhadores positivamente selecionados e com isso o aumento da dispersão entre os salários de migrantes e paulistas.

A Figura 4 mostra que sem migração de retorno a distribuição salarial dos migrantes estaria mais próxima a dos naturais, e que há uma lacuna entre os ganhos desses dois grupos. Na próxima seção podemos verificar se essa diferença se verifica em todos os níveis de capital humano e se a migração de retorno tem efeito sobre essa distância a partir de testes de heterogeneidade dos resultados apresentados.

Tabela 7 – Diferenças salariais entre migrantes e naturais de São Paulo - Cenário real e contrafactual

	Observáveis			Salário Total		
	Fatual (1)	Contrafactual (2)	Diferença (2)-(1)	Fatual (1)	Contrafactual (2)	Diferença (2)-(1)
1º decil	-0,2964	0,1067	0,4031	-0,1494	-1,0039	-0,8545
2º decil	0,1506	0,1603	0,0097	0,2113	0,1219	-0,0894
3º decil	0,1972	0,2010	0,0039	0,2355	0,1387	-0,0968
4º decil	0,2377	0,2374	-0,0003	0,2598	0,1713	-0,0885
5º decil	0,2806	0,2765	-0,0041	0,2970	0,2379	-0,0590
6º decil	0,3245	0,3163	-0,0083	0,3443	0,3138	-0,0304
7º decil	0,3739	0,3587	-0,0152	0,4014	0,4135	0,0120
8º decil	0,6110	0,5817	-0,0294	0,4629	0,5384	0,0755
9º decil	0,4494	-0,6349	-1,0843	0,7145	0,7049	-0,0096
Média	0,3084	0,2988	-0,0096	0,3084	0,2656	-0,0428

Fonte: Elaboração própria com base nos dados do Censo Demográfico de 2010. Nota: Os valores correspondem às diferenças entre os decis da distribuição salarial dos nativos e as distribuições reais e contrafatuais salariais dos migrantes. As predições de salários dos migrantes foram baseadas nos parâmetros estimados para o modelo bivariado do tipo cópula *Joe* rotacionada a 180°. Já as predições salariais dos não migrantes foram baseadas em um modelo aditivo generalizados com estimativa integrada de suavidade, mais detalhes em Wood e Wood (2015).

## 5.4 Efeitos da migração de retorno considerando diferentes níveis educacionais

O nível educacional dos migrantes com maior propensão a não remigrar se mostrou muito baixo. De acordo com a Tabela 4, há uma menor probabilidade de permanência para os níveis educacionais mais altos, comparados aos indivíduos sem instrução. Dessa forma, evidencia-se que a escolaridade pode afetar a decisão do migrante em se manter no destino. Portanto, para testar efeitos de heterogeneidade sobre as consequências da migração de retorno foram obtidas as distribuições reais e contrafatuais separando a amostra por nível de instrução. A distribuição total de salários, no cenário em que não há migração de retorno, foi obtida a partir do grupo de alta probabilidade agora condicionado aos diferentes níveis de educação considerados neste estudo, que serão comparadas às distribuições reais dos migrantes permanentes.

Tabela 8 – Decis de distribuições de salários dos migrantes com e sem migração de retorno por nível educacional

	Sem Instrução			Fundamental			Médio		
	(1) Fat.	(2) Cont.	(2) - (1)	(1) Fat.	(2) Cont.	(2) - (1)	(1) Fat.	(2) Cont.	(2) - (1).
Observáveis	$\hat{Y}_i$								
1º	1,1536	0,7506	-0,4031	1,4395	0,7758	-0,6638	1,4395	1,0336	-0,4059
2º	1,4775	1,4677	-0,0098	1,7121	1,5650	-0,1470	1,7121	1,6880	-0,0241
3º	1,5233	1,5200	-0,0033	1,7532	1,6085	-0,1447	1,7532	1,7414	-0,0118
4º	1,5598	1,5601	0,0003	1,7855	1,6464	-0,1391	1,7855	1,7789	-0,0066
5º	1,5918	1,5951	0,0033	1,8185	1,6851	-0,1334	1,8185	1,8166	-0,001
6º	1,6234	1,6290	0,0056	1,8528	1,7204	-0,1323	1,8528	1,8543	0,0015
7º	1,6584	1,6710	0,0126	1,8917	1,7654	-0,1263	1,8917	1,9007	0,0090
8º	1,7172	1,7392	0,0220	1,9460	1,8354	-0,1106	1,9460	1,9629	0,0169
9º	2,2931	2,9607	0,6676	2,4132	3,1168	0,7036	2,4132	3,1093	0,6961
Média	1,5937	1,6023	0,0086	1,8268	1,6922	-0,1347	1,8268	1,8227	-0,0041
Não Observáveis	$\hat{u}_i$								
1º	-0,9924	-0,7649	0,2275	-1,3585	-0,8412	0,5173	-1,1603	-0,8846	0,2756
2º	-0,4804	-0,3887	0,0917	-0,4756	-0,4541	0,0215	-0,5112	-0,5018	0,0094
3º	-0,3078	-0,2027	0,1051	-0,2997	-0,2488	0,0509	-0,3299	-0,3427	-0,0127
4º	-0,1716	-0,0741	0,0976	-0,1690	-0,1495	0,0195	-0,1919	-0,2225	-0,0306
5º	-0,0554	0,0140	0,0693	-0,0552	-0,0339	0,0213	-0,0658	-0,0886	-0,0229
6º	0,0663	0,1158	0,0494	0,0696	0,0564	-0,0133	0,0786	0,0041	-0,0744
7º	0,2128	0,2346	0,0218	0,2234	0,1636	-0,0598	0,2770	0,2168	-0,0603
8º	0,4644	0,4270	-0,0374	0,4955	0,4202	-0,0753	0,5615	0,3615	-0,2001
9º	1,9032	1,9032	0,0000	1,7710	1,7264	-0,0446	1,6786	1,4265	-0,2521
Média	0,0000	0,0534	0,0534	0,0000	0,0057	0,0057	0,0000	-0,0620	-0,0620
Total	$Y_i$								
1º	0,6991	0,9361	0,2371	0,3330	0,8599	0,5269	0,5312	0,8164	0,2852
2º	1,2111	1,3124	0,1013	1,2159	1,2470	0,0311	1,1803	1,1993	0,0190
3º	1,3836	1,4983	0,1147	1,3918	1,4523	0,0605	1,3615	1,3584	-0,0031
4º	1,5198	1,6270	0,1072	1,5225	1,5515	0,0290	1,4996	1,4786	-0,0210
5º	1,6361	1,7150	0,0789	1,6362	1,6671	0,0309	1,6257	1,6124	-0,0133
6º	1,7578	1,8169	0,0590	1,7611	1,7574	-0,0037	1,7700	1,7052	-0,0648
7º	1,9043	1,9357	0,0314	1,9148	1,8646	-0,0502	1,9685	1,9178	-0,0507
8º	2,1559	2,1281	-0,0278	2,1870	2,1213	-0,0657	2,2530	2,0625	-0,1905
9º	3,5946	3,6042	0,0096	3,4625	3,4274	-0,0351	3,3701	3,1276	-0,2425
Média	1,6915	1,7545	0,0630	1,6915	1,7068	0,0153	1,6915	1,6390	-0,0524
Diferenças Salariais									
10º - 90º	1,0788	0,9075	-0,1713	1,1056	0,9869	-0,1186	1,2292	0,9441	-0,2851
10º - 50º	0,4632	0,4317	-0,0315	0,4688	0,4664	-0,0024	0,5007	0,4614	-0,0393
50º - 90º	0,6157	0,4758	-0,1398	1,1056	0,5206	-0,5850	0,7285	0,4827	-0,2458

Fonte: Elaboração própria com base nos dados do Censo Demográfico de 2010

Nota 1: Nota 4: As predições de salários foram baseadas nos parâmetros estimados para o modelo bivariado do tipo cópula *Joe* rotacionada a 180°. Valores calculados condicionando as distribuições real e contrafactual a pessoas com ensino fundamental, médio e sem instrução (Ver Tabela 6). As diferenças salariais referem-se as diferenças nos percentis da distribuições salariais reais e contrafatuais.

A Tabela 8 é composta pelos decis das distribuições de salários-hora para os níveis: sem instrução, nível fundamental e médio<sup>3</sup>. As diferenças em observáveis

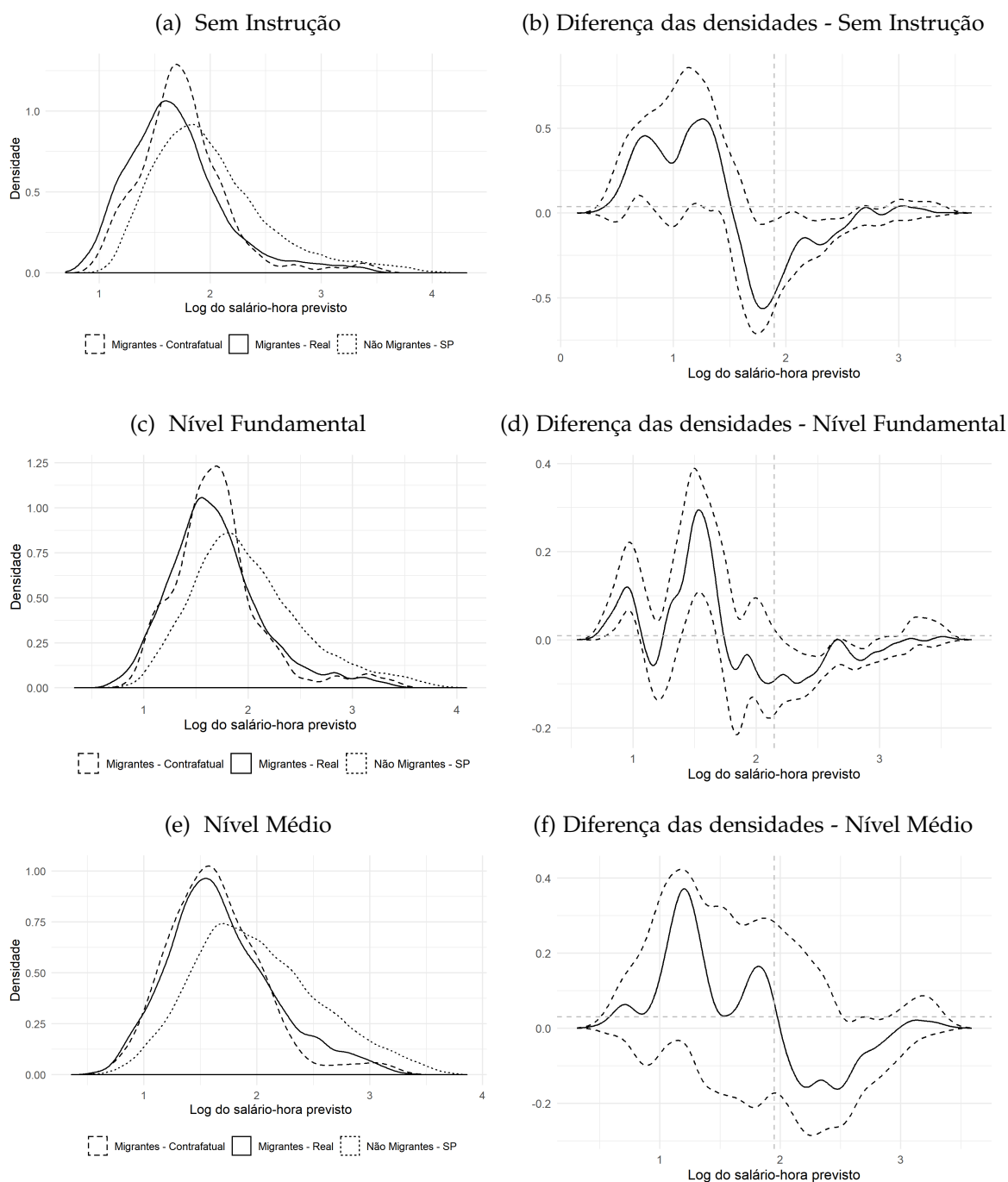
<sup>3</sup> Não foram identificados indivíduos com nível superior ou pós graduação no GAP o que impossibilita a construção da perspectiva contrafactual em características não observáveis nesses níveis. No entanto, são relatados os decis em observáveis para os dois níveis conjuntamente na Tabela ?? do apêndice

não são negligenciáveis, principalmente para o nível fundamental, que mostra um desfavorecimento para esse nível em termos de atributos mensuráveis. A diferença logarítmica ao longo dos decis é quase totalmente negativa, exceto pelo 9º decil, o que representa uma piora salarial de 13,47%, em média, no cenário em que não há migração de retorno. Considerando o salário total, no entanto, verificamos que há uma mudança significativa no efeito da remigração. Ao contar com as características inatas as diferenças entre as distribuições real e contrafactual dos trabalhadores com ensino fundamental é positiva até a mediana (5º decil), o que melhora, em média, os salários em 1,53%. Para os indivíduos sem instrução os resultados são muito próximos aos da amostra completa devido à maior presença de indivíduos nesse nível educacional na população migrante em São Paulo. Sem migração de retorno esse grupo estaria ganhando 6,3% a mais, em média, considerando o salário total. O nível médio, por sua vez, apresenta piora da média salarial considerando características inatas e mensuráveis, -0,41% e -5,24%, respectivamente. Esses resultados reforçam a conclusão geral de que indivíduos com baixa qualificação tendem a apresentar traços não observáveis superiores.

Na Figura 5 temos as densidades reais e contrafatuais nesses níveis educacionais. Para o nível sem instrução (a e b), como visto acima, há uma grande aproximação do resultado para amostra total, devido à grande proporção de indivíduos com esse nível de escolaridade. As densidades (b) e (c), para o nível fundamental, indicam um resultado semelhante, embora menos significativo, há um ligeiro deslocamento à direita da parte esquerda e topo da distribuição contrafactual. As densidades referentes ao ensino médio indicam um deslocamento à esquerda na metade inferior da distribuição, bem como menos migrantes na metade superior, mostrando que as diferenças entre as densidades para valores mais altos de rendimento são fortemente negativas. Destarte, para o nível médio o retorno representa uma piora salarial.

Nos grupos educacionais onde foi possível analisar isoladamente o efeito da migração de retorno sobre a desigualdade entre os migrantes em São Paulo foi possível concluir que há um aumento dos diferenciais salariais com migração de retorno. Em relação aos trabalhadores não migrantes, podemos concluir que em todos os grupos de capital humano considerados há uma lacuna favorável aos nativos. Sem migração de retorno esse hiato diminuiria ligeiramente para os trabalhadores no grupo sem instrução, para os trabalhadores com nível fundamental ou médio haveria uma distância ainda maior na metade superior das distribuições.

Figura 5 – Densidades reais e contrafatuais do salário-hora total para migrantes e paulistas por nível educacional



Fonte: Elaboração própria com base nos dados do Censo 2010.

Nota 1: A distribuição real foi construída a partir da amostra de migrantes residentes em SP condicionada aos três níveis de instrução supracitados. A distribuição contrafactual corresponde a distribuição dos salários totais no cenário em que não há migração de retorno condicionada aos mesmos níveis educacionais, calculada pela mudança dos resíduos no GAP pela média dos salários previstos dos migrantes caso todos tivessem permanecido.

Nota 2: O teste K-S indicou diferença entre a distribuição real e contrafactual significativa a 1% no nível sem instrução com estatística  $D = 0.4669$  e  $p\text{-valor} = 0,0000$ . Nível fundamental:  $D = 0.0670$  e  $p\text{-valor} = 0,1560$ . Nível médio:  $D = 0.0896$  e  $p\text{-valor} = 0.2450$ .

Nota 4: As predições de salários foram baseadas nos parâmetros estimados para o modelo bivariado do tipo cópula *Joe* rotacionada a  $180^\circ$ .

## 5.5 Análise de Robustez

Na seção anterior foram apresentados os resultados da análise principal a partir de um modelo flexível semiparamétrico. Nesta seção, verificamos a consistência desses resultados a partir de uma especificação onde não há restrição de exclusão e aplicação de teste de validade para variável instrumental. Ademais, também são registrados os resultados obtidos a partir da inclusão de mulheres na amostra.

### 5.5.1 Modelo sem restrição de exclusão e validade do instrumento

Muitas aplicações empíricas apoiam-se no uso de restrições de exclusões com o objetivo de auxiliar na correta identificação do modelo com viés de seleção amostral (TOOMET; HENNINGSEN et al., 2008). No modelo de Heckman (1979), por exemplo, além da premissa restritiva de normalidade conjunta das equações de seleção e salários, a correlação dos seus termos randômicos requer o uso de restrição de exclusão (variável instrumental) para garantir a correta identificação dos parâmetros estimados. Quando o modelo de determinação conjunta de remigração e salários (vide equações 4.1-4.2) tem sua distribuição conjunta baseada em cópula, não necessariamente precisamos de uma identificação por restrição de exclusão. Todavia, podemos obter estimativas mais consistentes quando ao menos um regressor a mais é incluído na equação de seleção (WOJTYS; MARRA; RADICE, 2016).

A variável filho nascido em São Paulo foi usada neste estudo entrando na equação de seleção e excluída da equação de salários (instrumento de exclusão), sob a premissa de que a convivência com filhos na região de destino cria um vínculo familiar que pode dificultar a migração de retorno (BIAVASCHI, 2016). A Figura 6 (e) mostra as densidades reais e contrafatuais para um modelo<sup>4</sup> de seleção amostral semelhante empregado na análise principal que não usa essa variável como restrição de exclusão<sup>5</sup>. Verifica-se que os resultados são consistentes com a análise principal. Para verificar a validade desse instrumento, no entanto, tendo em vista que os pressupostos requeridos ao usar essa estratégia de identificação, foram aplicados os testes propostos por Huber e Mellace (2014) e os métodos de Chen e Szroeter (2014) e Bennett et al. (2009) para testar estatisticamente restrições de desigualdade.

Huber e Mellace (2014) propõem testar a validade do instrumento e a separabilidade aditiva do termo de erro na equação de seleção conjuntamente<sup>6</sup>. A intuição do teste é que duas restrições de desigualdade surgem da identificação pontual e delimitação da distribuição dos indivíduos sempre selecionados (aqueles selecionados

<sup>4</sup> Na Tabela A3 do apêndice estão dispostos os valores do critério de informação AIC para o modelo sem restrição de exclusão. A cópula Joe rotacionada a 180° obteve o menor valor AIC.

<sup>5</sup> O mesmo vetor de variáveis foi usado nas duas equações 4.1 e 4.2.

<sup>6</sup> A restrição de exclusão é independente de  $u_i^*, \epsilon_i | X$ . Em termos de monotonicidade isso implica que o estado potencial de seleção de cada indivíduo aumenta ou diminui fracamente monotonicamente com o instrumento. Dessa forma, denotando por  $S(z)$  o estado potencial de seleção do instrumento, temos que  $\Pr(S(1) \geq S(0) | X) = 1$  (monotonicidade positiva) e  $\Pr(S(1) \leq S(0) | X) = 1$  (monotonicidade negativa) (HUBER; MELLACE, 2014).



independente do instrumento – *always takers*) na subpopulação selecionada que não recebe o instrumento, considerando que a população que recebe o instrumento inclui *compliers*, cuja seleção reage ao instrumento. O ponto identificado na ausência do instrumento deve estar dentro dos limites na presença dele, gerando duas restrições testáveis:

$$H_0 : \begin{pmatrix} E(Y|Z = 1, S = 1, Y \leq y_q) - E(Y|Z = 0, S = 1) \\ E(Y|Z = 0, S = 1) - E(Y|Z = 1, S = 1, Y \geq y_{1-q}) \end{pmatrix} \equiv \begin{pmatrix} \theta_1^m \\ \theta_2^m \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (5.1)$$

Onde  $q$  é a proporção de migrantes sempre selecionados na população mista de indivíduos com e sem filho nascido em São Paulo e  $y_q$  é o quantil da variável de resultado (logaritmo do salário-hora) condicionado à probabilidade  $q$ . Para testar essas restrições conjuntamente Huber e Mellace (2014) utilizam os métodos de Chen e Szroeter (2014). Foram também utilizados os métodos de Bennett et al. (2009) baseados em reamostragem por *bootstrap* para obtenção de p-valores robustos em amostras finitas<sup>7</sup>.

Os resultados dos testes das restrições em 5.1 estão dispostos na Tabela A7 do Apêndice. São reportados os resultados para várias subamostras condicionadas à diferentes subconjuntos das covariadas utilizadas neste estudo. A terceira coluna contém a porcentagem de *compliers*<sup>8</sup> em cada subamostra, a quarta coluna contém a distância padronizada, que indica o grau de violação das restrições de desigualdade em 5.1 (HUBER; MELLACE, 2014). A partir da quinta coluna são dispostos os p-valores obtidos a partir dos métodos de Chen e Szroeter (2014) e Bennett et al. (2009) com recentramento completo e parcial, respectivamente.

A partir dos resultados verificamos que a distância padronizada é negativa em grande parte das subamostras. Ademais, todas as estatísticas de p-valor são consideravelmente maiores que 10%. Assim posto, no geral, não há evidências para rejeitar a validade do instrumento utilizado neste estudo, especialmente nos casos em que o poder do teste é maior pela menor presença de *compliers*.

## 5.5.2 Análise incluindo mulheres

As mulheres representam 36,36% da amostra total caracterizada pelos recortes adotados na seção 4. Ao incluir esse grupo na amostra, no entanto, dois problemas podem surgir na análise. Primeiro, em muitos estudos sobre migração foi mostrado que as mulheres têm maiores chances de serem migrantes agregados (SIQUEIRA, 2006;

<sup>7</sup> Detalhes sobre algoritmos em Chen e Szroeter (2014) e Bennett et al. (2009)

<sup>8</sup> Huber e Mellace (2014) destacam que o poder do teste diminui quanto maior o percentual desse grupo.

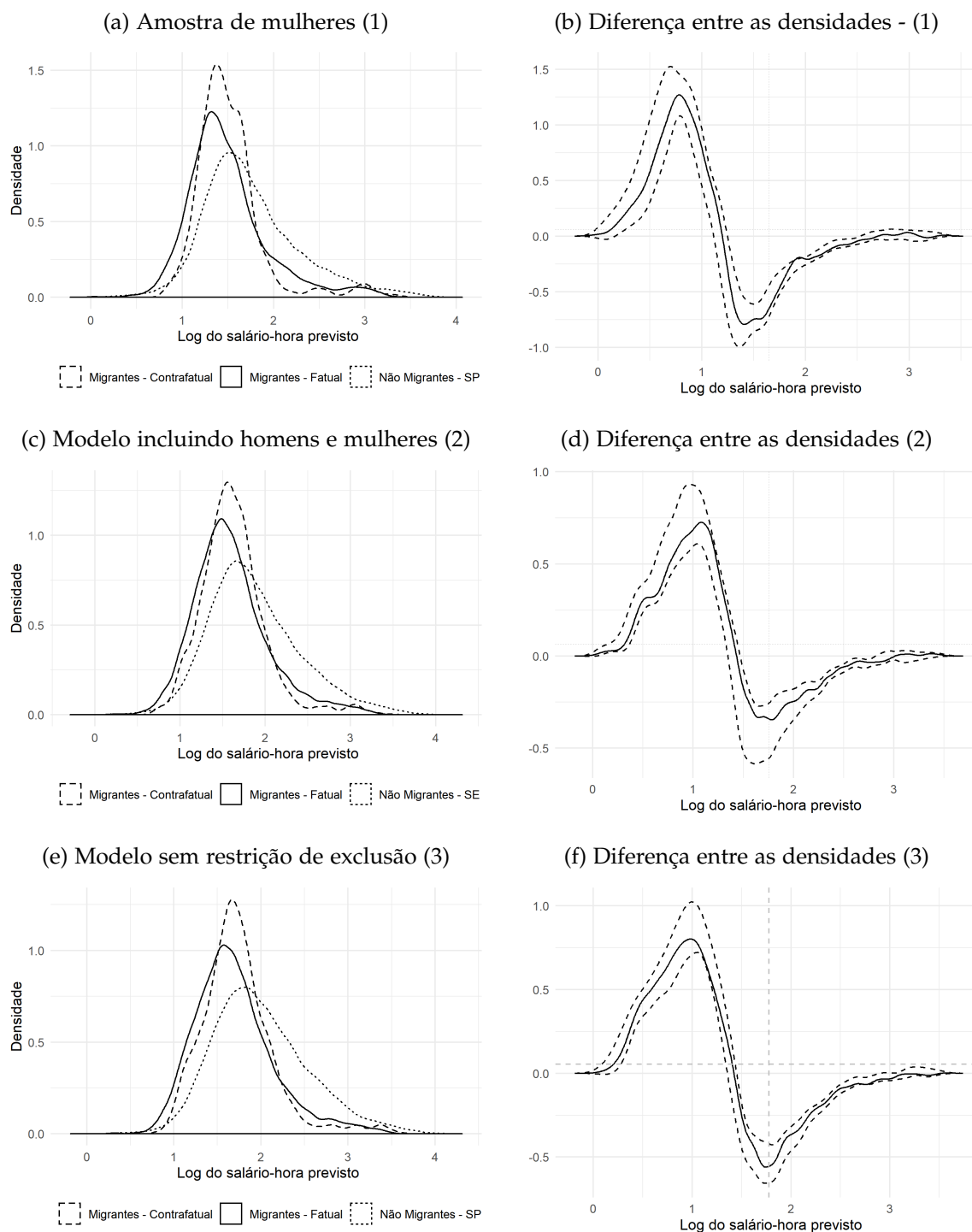


AVELINO, 2010; OLIVEIRA; JANNUZZI, 2016; AGUIAR, 2017). Segundo, a restrição de exclusão (residente com filho nascido em São Paulo) no modelo de seleção tem relação com a produtividade da mulher no mercado de trabalho. Há evidências de que a maternidade cria uma série de barreiras para a participação da mulher no mercado de trabalho, principalmente no curto prazo. Apesar disso, a participação das mulheres vem aumentando, enquanto as desigualdades e a problemática da jornada dupla persistem (PAZELLO; FERNANDES, 2004; COSTA, 2007; FONTOURA; GONZALEZ, 2009; QUEIROZ; ARAGÓN, 2015).

Assim posto, ao longo deste estudo a análise empírica foi realizada a partir de uma amostra composta apenas por trabalhadores homens. Essa estratégia foi implementada afim de possibilitar maior confiança na exogeneidade da variável de restrição de exclusão em relação à equação de resultados (rendimentos), isto é, ao reduzir risco de correlação entre cuidado com os filhos e produtividade do trabalho. No entanto, pode haver uma preocupação de que uma amostra que inclua as mulheres altere os resultados. Para verificar se os resultados são consistentes na presença de mulheres foram estimados dois modelos: o primeiro modelo considera apenas mulheres, enquanto o segundo inclui ambos homens e mulheres.

Considerando somente as mulheres, a Figura 6 (a) relata as densidades reais e contrafatuais do logaritmo do salário-hora. Verifica-se que os resultados são muito próximos aos da análise principal. Note-se que há um deslocamento à direita da distribuição de salários sem migração de retorno, principalmente no topo e metade e inferior, mais uma vez sugerindo que migrantes com rendimentos mais baixos e médios estariam auferindo maiores salários caso não houvesse migração de retorno. A Figura 6 (c) representa amostra em que são considerados homens e mulheres conjuntamente. Esses resultados são, mais uma vez, consistentes com as conclusões da amostra principal.

Figura 6 – Densidade reais e contrafatuais do salário-hora total para migrantes e paulistas - (1) Amostra de mulheres, (2) Modelo incluindo homens e mulheres e (3) Modelo sem restrição de exclusão



Fonte: Elaboração própria a partir dos dados do Censo Demográfico de 2010

Nota 1 : (2)(a) Amostra incluindo homens e mulheres. A distribuição real corresponde à distribuição dos salários totais dos migrantes permanentes (residentes em SP), o contrafactual corresponde a distribuição dos salários totais dos migrantes interestaduais caso todos pudessem no destino. Nota 2: (1)(b) e (2)(d) Intervalos de confiança representado pelas linhas pontilhadas, supondo um nível de 95% de confiança. A estimação dos erros-padrão foi feita usando a técnica de reamostragem *bootstrap*.

Nota 3: O teste *Kolmogorov - Smirnov* rejeitou-se a hipótese nula de que as densidades reais e contrafatuais sejam iguais, ao nível de 1% de significância em todos os modelos conforme os seguintes dados: (1)  $D = 0.0954$  e  $p\text{-valor} = 0,0000$ ; (2)  $D = 0.0941$  e  $p\text{-valor} = 0,0000$ ; (3)  $D = 0.0941$  e  $p\text{-valor} = 0,0000$

Nota 4: As previsões de salários foram baseadas nos parâmetros estimados para o modelo bivariado do tipo cópula *Joe* rotacionada a 180° em todos os modelos.

## 6 Considerações Finais

Conforme resultado teórico de [Borjas e Bratsberg \(1996\)](#), os efeitos do fenômeno da migração de retorno podem se rebater na população migrante que não optou pela remigração, uma vez que o fluxo de migração inicial tenha se autosselecionado em atributos produtivos não observados. Tal questão permanecia pouco explorada na literatura nacional.

Este estudo procurou preencher a lacuna reportada ao investigar empiricamente qual seriam os efeitos da migração de retorno sobre a distribuição de salários dos migrantes interestaduais que optaram pela permanência no estado de São Paulo. Tal estado, além de representar a maior economia dentre as demais unidades federativas do Brasil, tem recebido historicamente o maior contingente de migrantes, bem como se diferenciado como origem principal dos remigrados. Além de características mensuráveis como escolaridade, idade, entre outras, foram também consideradas características inatas na construção dessas distribuições, de forma a verificar se o impacto da migração de retorno sobre a população migrante pode ser explicado por questões intrínsecas ao trabalhador.

Para recuperar as distribuição contrafactual na presença de seleção na migração de retorno foram aplicados os métodos propostos por [Biavaschi \(2016\)](#). O procedimento para recuperar a distribuição de inobserváveis, conhecido como identificação no infinito, baseia-se no pressuposto de que o viés de seleção amostral desaparece no grupo indivíduos com alta probabilidade de permanecer em São Paulo. Ademais, a recuperação da distribuição total de salários também requer previsões consistentes a partir de parâmetros obtidos por um modelo de determinação conjunta de decisão de remigração e salários. A dependência entre as equações do referido modelo foi estimada a partir de um modelo bivariado semiparamétrico baseado em cópulas, que flexibiliza premissas sobre a distribuição estatística dos dados e identificação causal ([RADICE; MARRA; WOJTYŚ, 2016](#)).

As estimações dos determinantes da escolha de permanência identificaram que os laços criados na região de destino podem ter forte influência sobre a decisão de remigrar. Aqueles casados com cônjuges nascidos em São Paulo e/ou com filhos nascidos no estado mostraram maior propensão a se tornar migrantes permanentes. Os achados também sugerem que indivíduos com níveis mais elevados de instrução tendem a ser migrantes de retorno. Ademais, chefes de família, migrantes com filhos menores que 14 anos, autônomos, empregadores, beneficiários de programa de transferência de renda do governo e migrantes oriundos das regiões Sudeste e Sul são

mais propensos ao retorno. Do ponto de vista dos rendimentos, verificou-se que os migrantes em São Paulo com maiores níveis de instrução, brancos, chefes de família sem filhos, que não sejam trabalhadores informais e não beneficiários de programas sociais auferem salários maiores. Quanto à região de origem, migrantes com origem no Nordeste recebem salários inferiores em relação àqueles provenientes do Sudeste.

As densidades estimadas mostram que, caso não houvesse migração de retorno, mais massa de migrantes apareceria na metade superior da distribuição, além de haver um deslocamento à direita da parte esquerda da distribuição, melhorando o salário médio da população migrante. Na cauda superior da distribuição, para aqueles indivíduos com salários mais altos, a ausência de migração de retorno representa uma piora salarial. Esse balanceamento do efeito entre os extremos da distribuição de salários é responsável pela queda na desigualdade salarial entre os migrantes no cenário contrafactual, sugerindo que a migração seletiva de retorno piora as desigualdades salariais entre o próprio grupo de migrantes. Sem migração de retorno os migrantes estariam recebendo 4,28% a mais na média e a desigualdade entre os percentis 10 e 90 diminuiria 18,1%. Esses resultados se estendem na análise por grupos educacionais, mostrando que o impacto da remigração é maior e mais positivo para o grupo de trabalhadores menos qualificados, considerando características inatas. Também foi verificado que resultados permanecem estáveis quando verificados em diferentes amostras e técnicas.

Outra conclusão importante a ser ressaltada é a disparidade de rendimentos entre migrantes e não migrantes em todos os níveis de capital humano. A desigualdade salarial entre esses grupos, no entanto, seria menor na ausência de migração de retorno, devido à melhora na média salarial dos migrantes. Quando controlamos o impacto da remigração entre grupos de diferentes níveis de instrução, os achados relevaram efeitos heterogêneos para toda população migrante. Ou seja, a média salarial dos trabalhadores de baixa instrução aumenta sem migração de retorno, enquanto para migrantes com maior instrução, o salário médio diminui.

Portanto, os achados deste estudo revelam que o fluxo de migração interestadual dirigido ao maior polo econômico do Brasil parece ser formado por trabalhadores negativamente selecionados, não apenas em atributos produtivos observáveis (baixa instrução), mas também em características produtivas não observadas. O último fato é reforçado pelas evidências de que os migrantes retornados têm características produtivas não observadas melhores quando comparadas aos migrantes permanentes, as quais, em geral, aumentariam o salário médio da população migrante e reduziria a desigualdade caso não houvesse migração de retorno.

Como sugestão para estudos futuros está a análise do impacto da migração de retorno sobre as regiões de origem dos migrantes em termos de desigualdade salarial e

inserção do retornado no mercado de trabalho local. Além disso, sugere-se estender a investigação para outras regiões de absorção de migrantes, como a região Centro-Oeste, tendo em vista as mudanças nas tendências migratórias observadas e discutidas na literatura nacional.

# Referências

- AGUIAR, M. A. S. d. Autosseleção e impacto da migração de retorno sobre a distribuição salarial: análise para os migrantes da região nordeste. 2017.
- AMBROSINI, J. W. et al. *The selection of migrants and returnees: Evidence from Romania and implications*. [S.l.]: World Bank, 2010.
- ANDREWS, D. W.; SCHAFGANS, M. M. Semiparametric estimation of the intercept of a sample selection model. *The Review of Economic Studies*, Wiley-Blackwell, v. 65, n. 3, p. 497–517, 1998.
- ASSIS, R. d.; COSTA, E. M.; MARIANO, J. Impacto da migração de não naturais e da migração de retorno sobre a distribuição de renda dos estados da bahia e de são paulo: um olhar sobre a inserção desses indivíduos no mercado de trabalho local. 40<sup>o</sup> *Encontro Nacional de Economia*, 2012.
- AVELINO, R. R. Self-selection and the impact of migration on earnings. *Brazilian Review of Econometrics*, v. 30, n. 1, p. 69–89, 2010.
- BAENINGER, R. Rotatividade migratória: um novo olhar para as migrações no século xxi. *Anais*, p. 1–21, 2016.
- BAENINGER, R. A. et al. Região, metrópole e interior: espaços ganhadores e espaços perdedores nas migrações recentes-brasil, 1980-1996. [sn], 1999.
- BATISTA, N. N. F.; CACCIAMALI, M. C. Diferencial de salários entre homens e mulheres segundo a condição de migração. *Revista Brasileira de Estudos de População*, SciELO Brasil, v. 26, n. 1, p. 97–115, 2009.
- BENNETT, C. J. et al. Consistent and asymptotically unbiased minp tests of multiple inequality moment restrictions. *Vanderbilt University Department of Economics Working Papers*, Vanderbilt University Department of Economics, v. 908, 2009.
- BERTOLI, S.; MORAGA, J. F.-H.; ORTEGA, F. Crossing the border: Self-selection, earnings and individual migration decisions. *Journal of Development Economics*, Elsevier, v. 101, p. 75–91, 2013.
- BIAVASCHI, C. Recovering the counterfactual wage distribution with selective return migration. *Labour Economics*, Elsevier, v. 38, p. 59–80, 2016.
- BORJAS, G. J. *Self-selection and the earnings of immigrants*. [S.l.]: National Bureau of Economic Research Cambridge, Mass., USA, 1987.
- BORJAS, G. J.; BRATSBERG, B. *Who leaves? The outmigration of the foreign-born*. [S.l.], 1996.
- BRECHMANN, E.; SCHEPSMEIER, U. Cdvine: Modeling dependence with c-and d-vine copulas in r. *Journal of Statistical Software*, v. 52, n. 3, p. 1–27, 2013.

- BRITO, A. S. Efeitos esperados pela teoria econômica de políticas de transferência de renda sobre o mercado de trabalho. 2011.
- BRITO, F. Brasil, final de século: a transição para um novo padrão migratório? *Anais*, p. 1–44, 2016.
- CAMERON, A. C.; TRIVEDI, P. K. *Microeconometrics: methods and applications*. [S.l.]: Cambridge university press, 2005.
- CASSARINO, J.-P. Theorising return migration: The conceptual approach to return migrants revisited. *International Journal on Multicultural Societies (IJMS)*, v. 6, n. 2, p. 253–279, 2004.
- CATTANEO, C. et al. The self-selection in the migration process: What can we learn? Citeseer, 2007.
- CHAMBERLAIN, G. Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, Elsevier, v. 32, n. 2, p. 189–218, 1986.
- CHEN, L.-Y.; SZROETER, J. Testing multiple inequality hypotheses: a smoothed indicator approach. *Journal of econometrics*, Elsevier, v. 178, p. 678–693, 2014.
- CHIQUEL, D.; HANSON, G. H. International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. *Journal of Political Economy*, The University of Chicago Press, v. 113, n. 2, p. 239–281, 2005.
- CHISWICK, B. Are immigrants favorably self-selected? *American Economic Review*, v. 89, n. 2, p. 181–185, 1999.
- CHZHEN, Y.; MUMFORD, K. Gender gaps across the earnings distribution for full-time employees in Britain: Allowing for sample selection. *Labour Economics*, Elsevier, v. 18, n. 6, p. 837–844, 2011.
- COHEN, Y.; HABERFELD, Y. Self-selection and return migration: Israeli-born Jews returning home from the United States during the 1980s. *Population Studies*, Taylor & Francis, v. 55, n. 1, p. 79–91, 2001.
- CONSTANT, A.; MASSEY, D. S. Return migration by German guestworkers: Neoclassical versus new economic theories. *International Migration*, Wiley Online Library, v. 40, n. 4, p. 5–38, 2002.
- COSTA, J. *Determinantes da participação feminina no mercado de trabalho brasileiro*. 2007. Tese (Doutorado) — Tese (Mestrado)—Universidade de Brasília, Brasília, 2007.
- COULON, A. D.; PIRACHA, M. Self-selection and the performance of return migrants: the source country perspective. *Journal of Population Economics*, Springer, v. 18, n. 4, p. 779–807, 2005.
- CUNHA, A. S. da. Migração de retorno num contexto de crises, mudanças e novos desafios. *Anais*, p. 1–36, 2016.
- CUNHA, J. M. P. Redistribuição espacial da população: tendências e trajetória. *São Paulo em perspectiva*, SciELO Brasil, v. 17, n. 3-4, p. 218–233, 2003.

DINARDO, J.; FORTIN, N. M.; LEMIEUX, T. *Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach*. [S.l.], 1995.

DUSTMANN, C. Return migration, wage differentials, and the optimal migration duration. *European Economic Review*, Elsevier, v. 47, n. 2, p. 353–369, 2003.

DUSTMANN, C.; KIRCHKAMP, O. The optimal migration duration and activity choice after re-migration. *Journal of development economics*, Elsevier, v. 67, n. 2, p. 351–372, 2002.

FONTOURA, N. d. O.; GONZALEZ, R. Aumento da participação de mulheres no mercado de trabalho: mudança ou reprodução da desigualdade? Instituto de Pesquisa Econômica Aplicada (Ipea), 2009.

FREGUGLIA, R. d. S.; PROCÓPIO, T. S. Efeitos da mudança de emprego e da migração interestadual sobre os salários no brasil formal: evidências a partir de dados em painel. Instituto de Pesquisa Econômica Aplicada (Ipea), 2013.

GAMA, L. C. D.; MACHADO, A. F. Migração e rendimentos no brasil: análise dos fatores associados no período intercensitário 2000-2010. *estudos avançados*, SciELO Brasil, v. 28, n. 81, p. 155–174, 2014.

HECKMAN, J. Varieties of selection bias. *The American Economic Review*, American Economic Association, v. 80, n. 2, p. 313, 1990.

HECKMAN, J. J. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, JSTOR, p. 153–161, 1979.

HUBER, M.; MELLACE, G. Testing exclusion restrictions and additive separability in sample selection models. *Empirical Economics*, Springer, v. 47, n. 1, p. 75–92, 2014.

IBGE, I. C. Resultados gerais da amostra. *Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística (IBGE)*, 2012.

JÚNIOR, E. d. R. d. S. et al. Migração, seleção e diferenças regionais de renda no brasil. Instituto de Pesquisa Econômica Aplicada (Ipea), 2005.

JUSTO, W. R.; FERREIRA, R. d. A. Migração interestadual no brasil: perfil do retornado: evidências para o período de 1998-2008. *XVIII Encontro Nacional de Estudos Populacionais*, 2012.

LEE, L.-F. Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables. *International economic review*, JSTOR, p. 415–433, 1978.

LEE, L.-F. Generalized econometric models with selectivity. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 507–512, 1983.

LIMA, A. C. d. C.; SIMÕES, R.; HERMETO, A. M. Privação relativa e deslocamentos da mão de obra no brasil entre 1980 e 2010: evolução das interações entre pobreza, desigualdade de renda e migração. Instituto de Pesquisa Econômica Aplicada (Ipea), 2015.



- LIU, E. et al. Maternal full-time employment and overweight children: Parametric, semi-parametric, and non-parametric assessment. *Journal of Econometrics*, Elsevier, v. 152, n. 1, p. 61–69, 2009.
- MARRA, G.; RADICE, R. Penalised regression splines: theory and application to medical research. *Statistical Methods in Medical Research*, Sage Publications Sage UK: London, England, v. 19, n. 2, p. 107–125, 2010.
- MARRA, G.; WYSZYNSKI, K. Semi-parametric copula sample selection models for count responses. *Computational Statistics & Data Analysis*, Elsevier, v. 104, p. 110–129, 2016.
- MASSEY, D. S.; GOLDRING, L.; DURAND, J. Continuities in transnational migration: An analysis of nineteen mexican communities. *American journal of Sociology*, University of Chicago Press, v. 99, n. 6, p. 1492–1533, 1994.
- MAYR, K.; PERI, G. *Return migration as a channel of brain gain*. [S.l.], 2008.
- MCKENZIE, D.; RAPOPORT, H. Self-selection patterns in mexico-us migration: the role of migration networks. *The Review of Economics and Statistics*, MIT Press, v. 92, n. 4, p. 811–821, 2010.
- MINCER, J. Family migration decisions. *Journal of political Economy*, The University of Chicago Press, v. 86, n. 5, p. 749–773, 1978.
- NETO, R. d. M. S.; MAGALHÃES, A. M. O progresso econômico do migrante em são paulo: evidências a partir dos censos demográficos de 1991 e 2000. *III Encontro da Associação Brasileira de Estudos Regionais-ABER*, 2004.
- NETTO JÚNIOR, J. L. et al. *Migrações e acumulação de capital humano: uma análise do período de 1950-2000*. [S.l.], 2003.
- NIKOLKA, T. Family return migration. Kiel, Hamburg: ZBW-Leibniz-Informationszentrum Wirtschaft, 2018.
- OLIVEIRA, A. d.; ERVATTI, L. R.; O'NEILL, M. O panorama dos deslocamentos populacionais no brasil: Pnads e censos demográficos. *Reflexões sobre os deslocamentos populacionais no Brasil*, Instituto Brasileiro de Geografia e Estatística-IBGE, Estudos e Análises . . . , v. 1, 2011.
- OLIVEIRA, K. F. de; JANNUZZI, P. de M. Motivos para migração no brasil: padrões etários, por sexo e origem/destino. *Anais*, p. 1–13, 2016.
- PACHECO, C. A. Evolução recente da urbanização e da questão regional no brasil: implicações econômicas para a dinâmica demográfica. In: *IV Conferencia Latinoamericana de población: la transición demográfica en América Latina y el Caribe*. [S.l.: s.n.], 1993. p. 41–56.
- PAZELLO, E. T.; FERNANDES, R. A maternidade e a mulher no mercado de trabalho: diferença de comportamento entre mulheres que têm e mulheres que não têm filhos. *ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO EM ECONOMIA*, v. 31, 2004.

- QUEIROZ, V. d. S.; ARAGÓN, J. A. O. Alocação de tempo em trabalho pelas mulheres brasileiras. *Estudos Econômicos (São Paulo)*, SciELO Brasil, v. 45, n. 4, p. 787–819, 2015.
- RADICE, R.; MARRA, G.; WOJTYŚ, M. Copula regression spline models for binary outcomes. *Statistics and Computing*, Springer, v. 26, n. 5, p. 981–995, 2016.
- RAMALHO, H. M. d. B.; FIGUEIREDO, E.; NETTO JÚNIOR, J. L. d. Determinantes das migrações interestaduais no Brasil: evidências a partir de um modelo gravitacional. Instituto de Pesquisa Econômica Aplicada (Ipea), 2016.
- RAMALHO, H. M. d. B.; QUEIROZ, V. d. S. Migração interestadual de retorno e autosseleção: evidências para o Brasil. 2011.
- SACHSIDA, A. et al. *Perfil do migrante brasileiro*. [S.l.], 2009.
- SANTOS, C. A. R. *Migração e distribuição regional de renda no Brasil*. Tese (Doutorado), 2007.
- SCHULTZ, T. W. Investment in human capital. *The American economic review*, JSTOR, p. 1–17, 1961.
- SHEN, C. Determinants of health care decisions: insurance, utilization, and expenditures. *Review of Economics and Statistics*, MIT Press, v. 95, n. 1, p. 142–153, 2013.
- SILVERMAN, B. W. *Density estimation for statistics and data analysis*. [S.l.]: Routledge, 2018.
- SIQUEIRA, L. B. O. d. Uma análise do fluxo migratório brasileiro: migração para regiões pobres e migração de retorno. Universidade Federal de Pernambuco, 2006.
- SJAASTAD, L. A. The costs and returns of human migration. *Journal of political Economy*, The University of Chicago Press, v. 70, n. 5, Part 2, p. 80–93, 1962.
- TEIXEIRA, C. G. Efeitos da transferência de renda na oferta de trabalho. Instituto de Pesquisa Econômica Aplicada (Ipea), 2011.
- TODARO, M. P. A model of labor migration and urban unemployment in less developed countries. *The American economic review*, JSTOR, v. 59, n. 1, p. 138–148, 1969.
- TOOMET, O.; HENNINGSEN, A. et al. Sample selection models in R: Package sampleselection. *Journal of statistical software*, v. 27, n. 7, p. 1–23, 2008.
- WOJTYŚ, M.; MARRA, G.; RADICE, R. Copula regression spline sample selection models: the R package semiparsamplesel. *Journal of Statistical Software*, v. 71, n. 6, 2016.
- WOOD, S.; WOOD, M. S. Package ‘mgcv’. *R package version*, v. 1, p. 29, 2015.

## 7 Apêndice

Tabela A1 – Variáveis socioeconômicas, educacionais e demográficas que foram utilizadas nos modelos de seleção e salários

Variáveis Pessoais	Descrição
Idade	Em anos completos
Idade2	Idade ao quadrado
Não Branco	Variável binária: 1 indivíduo que declarou cor ou raça diferente da branca ; 0 indivíduo declarou raça ou cor branca
Nível de Instrução	Categórica: define se o indivíduo não possui instrução ou tem o fundamental incompleto, se possui o fundamental completo ou médio incompleto, se possui o médio completo ou superior incompleto, se possui o superior completo ou se possui pós graduação.
Variáveis familiares	
Chefe	Variável binária: 1 posição de responsável pelo domicílio; 0 posição de cônjuge no domicílio
Chefe e vive com cônjuge	variável Binária: 1 chefe vive com cônjuge; 0 caso contrário
Chefe e vive com cônjuge nascido em SP	Variável binária: 1 chefe vive com cônjuge natural de São Paulo ; 0 caso contrário
Filho	Variável binária: 1 indivíduo possui filho com idade até 14 anos; 0 caso contrário
Filho nascido em SP	Variável binária: 1 indivíduo com filho nascido em SP com idade até 14 anos; 0 caso contrário
Total de pessoas no domicílio	- Número de pessoas morando no domicílio em 31 julho de 2010
Número de pessoas no domicílio2	Numero de pessoas no domicílio ao quadrado
Emprego e Ocupação	
Rendimento	Rendimento no trabalho principal
Salário-Hora	Rendimento do trabalho principal dividido pela média de horas trabalhadas por mês
Empregado com carteira de trabalho assinada	Variável binária: 1 empregado com carteira de trabalho assinada; 1 0 caso contrário
Autônomo	Variável binária: 1 trabalhador autônomo/conta-própria; 0 caso contrário
Empregador	Variável binária: 1 empregador; 0 caso contrário
Programa Social	Variável Binária: 1 Tinha rendimento de programa social ou transferência ( Bolsa Família ou Programa de erradicação do trabalho infantil - PETI) , 0 c.c
Residência	
Urbana	Variável binária: 1 reside na zona urbana; 0 caso contrário
Norte	Variável Binária: 1 se o indivíduo nasceu na região Norte, 0 caso contrário
Nordeste	Variável Binária: 1 se o indivíduo nasceu na região Nordeste, 0 caso contrário
Sudeste	Variável Binária: 1 se o indivíduo nasceu na região Sudeste, 0 caso contrário
Sul	variável Binária: 1 se o indivíduo nasceu na região Sul, 0 caso contrário
Centro-oeste	Variável Binária: 1 se o indivíduo nasceu na região centro-Oeste, 0 caso contrário

Fonte: Elaboração própria com base nos dados do Censo Demográfico de 2010.

Tabela A2 – Características das famílias de cópulas selecionadas

Cópula	Forma Funcional $C(u, v; \theta)$	Parâmetro de Dependência
Gaussiana	$\Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$	$\theta \in [-1, 1]$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$\theta \in (0, \infty)$
Joe	$1 - [(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta]^{1/\theta}$	$\theta \in (1, \infty)$
Frank	$-\theta^{-1} \log [1 + (e^{-\theta u} - 1)(e^{-\theta v} - 1) / (e^{-\theta} - 1)]$	$\theta \in \mathbb{R} \setminus \{0\}$
Gumbel	$\exp \left\{ - [(-\log u)^\theta + (-\log v)^\theta]^{1/\theta} \right\}$	$\theta \in [1, \infty)$
FGM	$uv[1 + \theta(1-u)(1-v)]$	$\theta \in [-1, 1]$
AMH	$uv/[1 - \theta(1-u)(1-v)]$	$\theta \in [-1, 1]$

Fonte: Wojtys, Marra e Radice (2016)

Nota:  $\Phi_2(\cdot, \cdot; \theta)$  denota a função de distribuição cumulativa de uma distribuição normal bivariada padrão com coeficiente de correlação  $\theta$ .  $\Phi^{-1}$  é a inversa da função acumulada normal padrão;  $u$  e  $v$  são distribuições marginais.

Tabela A3 – Critério de informação AIC: seleção do modelo segundo a função cópula - modelo principal

	(1)	(2)	(3)	(4)	(5)	(6)
Cópula	AIC	AIC	AIC	AIC	AIC	AIC
Joe 180	53.146,41	52.847,83	51.844,25	51.722,11	50.539,97	50.585,55
Clayton	53.162,08	52.865,16	51.880,59	51.764,07	50.590,01	50.630,49
Gumbel 180	53.189,85	52.900,95	51.932,52	51.816,20	-	-
Joe 270	53.840,87	53.476,69	52.211,39	51.938,38	50.712,47	50.980,07
Clayton 90	53.848,59	53.484,39	52.217,78	51.947,00	50.721,28	50.987,17
FGM	53.948,48	53.578,58	52.292,37	52.045,95	50.793,07	51.015,71
Frank	53.953,50	53.582,43	52.301,28	52.055,44	50.800,71	51.015,71
AMH	54.044,23	53.671,79	52.363,89	52.131,28	50.853,71	50.667,68
Gumbel 270	54.097,69	53.722,50	52.385,52	52.144,50	50.871,23	51.065,31
Gumbel	53.279,11	53.020,23	52.248,23	52.156,82	-	-
Clayton 180	53.361,47	53.153,02	52.393,77	52.184,57	50.879,47	51.010,45
Normal	54.140,85	53.766,52	52.430,88	52.211,60	-	-
Gumbel 90	54.140,87	53.766,52	52.430,88	52.213,87	-	-
Clayton 270	54.140,87	53.766,52	52.430,88	52.213,87	50.914,52	-
Joe 90	54.140,87	53.766,52	52.430,88	52.213,87	50.914,52	51.069,03
Joe	53.362,00	53.153,72	52.563,34	52.499,97	51.558,31	51.000,27
Especificação						
Variáveis Pessoais	sim	sim	sim	sim	sim	sim
Variáveis Familiares	não	sim	sim	sim	sim	sim
Emprego e Ocupação	não	não	sim	sim	sim	sim
Variáveis de Residência	não	não	não	sim	sim	sim
Restrição de exclusão	sim	sim	sim	sim	sim	não
Splines	não	não	não	não	sim	sim

Fonte: Elaboração própria a partir dos microdados do Censo Demográfico de 2010.

Nota 1 : Amostra de trabalhadores homens entre 18 e 74 anos em atividade remuneratória. Os modelos incluem duas partes: equação de seleção e equação de salários. Variáveis Pessoais: idade, idade ao quadrado, raça, e faixas de instrução e ainda uma variável binária indicando se o indivíduo é chefe de família. Variáveis familiares: variáveis binárias para filho nascido em SP (restrição de exclusão), filho menor de 14 anos, cônjuge nascido em SP, variável binária para chefe casado, quantitativo do total pessoas no domicílio, e número de pessoas no domicílio ao quadrado. Variáveis relacionadas a emprego e ocupação: variável binária para empregado com carteira assinada, *dummies* para condição de ocupação (autônomo e empregador) e rendimento de programa social. Variáveis de residência: setor do domicílio, variáveis binárias por região de origem do migrante. O modelo (5) utiliza, além dos componentes paramétricos, funções suavizadas (*splines*) das variáveis idade e número de pessoas no domicílio afim de capturar relações não lineares mais complexas. A tabela está ordenada de acordo com a coluna (4)

Tabela A4 – Critério de informação AIC: seleção do modelo segundo a função cópula - amostra de mulheres

	(1)	(2)	(3)	(4)	(5)
Cópula	AIC	AIC	AIC	AIC	AIC
Joe 180	29.202,88	29.074,95	28.703,72	28.357,38	28.076,84
Clayton	29.210,56	29.083,12	28.717,38	28.376,64	28.095,47
Gumbel 180	29.189,65	29.064,86	28.719,93	28.407,68	-
Gumbel	29.129,48	29.016,38	28.721,97	28.499,37	28.199,83
Clayton 180	29.082,99	28.990,19	28.730,71	28.584,43	28.290,23
Joe	29.083,33	28.990,59	28.731,01	28.584,92	28.290,62
Joe 270	29.972,50	29.785,79	29.169,97	28.591,62	28.375,03
Clayton 90	29.981,70	29.795,13	29.175,74	28.598,02	28.381,36
Frank	29.081,99	29.911,02	29.249,79	28.668,03	28.438,24
FGM	30.155,40	29.974,60	29.280,39	28.692,86	28.465,01
AMH	29.694,35	30.041,71	29.340,51	28.750,02	28.511,39
Gumbel 270	30.296,74	30.110,95	29.377,79	28.766,52	28.534,83
Normal	29.185,22	30.164,89	29.421,53	28.822,35	28.574,46
Clayton 270	30.341,15	30.165,24	29.421,86	28.825,59	28.576,79
Joe 90	30.341,15	30.165,24	29.421,86	28.825,59	28.576,79
Gumbel 90	30.341,15	30.165,24	29.421,86	-	-

Fonte: Elaboração própria a partir dos microdados do Censo Demográfico de 2010.

Nota 1 : A amostra inclui apenas migrantes mulheres entre 18 e 70 anos em atividade remuneratória. Os modelos incluem duas partes: equação de seleção e equação de salários. Variáveis Pessoais: idade, idade ao quadrado, raça, e faixas de instrução e ainda uma variável binária indicando se o indivíduo é chefe de família. Variáveis familiares: variáveis binárias para filho nascido em SP (restrição de exclusão), filho menor de 14 anos, cônjuge nascido em SP, variável binária para chefe casado, quantitativo do total pessoas no domicílio, e número de pessoas no domicílio ao quadrado. Variáveis relacionadas a emprego e ocupação: variável binária para empregado com carteira assinada, *dummies* para condição de ocupação (autônomo e empregador) e rendimento de programa social. Variáveis de residência: setor do domicílio, variáveis binárias por região de origem do migrante. O modelo (5) utiliza, além dos componentes paramétricos, funções suavizadas (*splines*) das variáveis idade e número de pessoas no domicílio afim de capturar relações não lineares mais complexas. A tabela está ordenada de acordo com a coluna (4)

Tabela A5 – Critério de informação AIC: seleção do modelo segundo a função cópula - amostra incluindo mulheres

	(1)	(2)	(3)	(4)	(5)
Cópula	AIC	AIC	AIC	AIC	AIC
Joe 180	81.535,41	81.097,57	79.622,11	78.405,91	77.553,98
Clayton	81.564,68	81.129,98	79.686,82	78.492,57	77.634,87
Gumbel 180	81.582,27	81.160,21	79.735,53	78.537,22	-
Joe 270	82.467,39	81.924,30	80.078,74	78.542,98	77.762,21
Clayton 90	82.482,62	81.939,60	80.089,00	78.560,03	77.776,09
Frank	82.667,55	82.118,67	80.221,47	78.724,40	77.895,77
FGM	82.685,48	82.141,71	80.217,65	78.728,25	77.895,67
Gumbel 270	82.937,43	82.388,58	80.370,36	78.835,90	78.014,20
AMH	82.838,12	82.294,30	80.335,98	78.837,48	77.991,39
Clayton 180	81.841,49	81.528,08	80.400,77	78.905,96	78.037,50
Normal	83.013,04	82.475,18	80.452,81	78.950,21	78.089,99
Gumbel 90	83.013,06	82.475,57	80.453,22	78.954,73	-
Clauton 270	83.013,06	82.475,57	80.453,22	78.954,73	78.091,52
Joe 90	83.013,06	82.475,57	80.453,22	78.954,73	78.090,43
Gumbel	81.676,62	81.298,44	80.181,12	79.202,51	-
Joe	81.845,43	81.532,52	80.658,31	79.939,02	79.045,00

Fonte: Elaboração própria a partir dos microdados do Censo Demográfico de 2010.

Nota 1 : Amostra de homens e mulheres entre 18 e 70 anos em atividade remuneratória. Os modelos incluem duas partes: equação de seleção e equação de salários. Variáveis Pessoais: idade, idade ao quadrado, raça, e faixas de instrução e ainda uma variável binária indicando se o indivíduo é chefe de família. Variáveis familiares: variáveis binárias para filho nascido em SP (restrição de exclusão), filho menor de 14 anos, cônjuge nascido em SP, variável binária para chefe casado, quantitativo do total pessoas no domicílio, e número de pessoas no domicílio ao quadrado. Variáveis relacionadas a emprego e ocupação: variável binária para empregado com carteira assinada, *dummies* para condição de ocupação (autônomo e empregador) e rendimento de programa social. Variáveis de residência: setor do domicílio, variáveis binárias por região de origem do migrante. O modelo (5) utiliza, além dos componentes paramétricos, funções suavizadas (*splines*) das variáveis idade e número de pessoas no domicílio afim de capturar relações não lineares mais complexas. A tabela está ordenada de acordo coma coluna (4)

Tabela A6 – Decis de distribuições de salários dos migrantes com e sem migração de retorno - nível de instrução superior e pós graduação

	Superior e Pós Graduação		
	(1) Fatural	(2) Contrafatural	(2) - (1) Diferença
Observáveis			
1º decil	2,2192	1,8098	-0,4094
2º decil	2,4075	2,3690	-0,0385
3º decil	2,4473	2,4309	-0,0164
4º decil	2,4768	2,4719	-0,0049
5º decil	2,5088	2,5088	0,0000
6º decil	2,5431	2,5498	0,0067
7º decil	2,5735	2,5985	0,0250
8º decil	2,6522	2,7131	0,0609
9º decil	3,0303	4,1146	1,0843
Média	2,5273	2,5351	0,0078
Amostra Total	738	990	-

Fonte: Elaboração própria com base nos dados do Censo 2010.

Nota 1: Valores calculados condicionando as distribuições real e contrafatural à pessoas com ensino superior ou pós graduação. Como o grupo de alta probabilidade não possui indivíduos com esse nível de instrução só é possível verificar os resultados em observáveis. (Ver Tabela 6)

Nota 2: As predições de salários foram baseadas nos parâmetros estimados para o modelo bivariado do tipo cópula *Joe* rotacionada a 180°

Tabela A7 – Teste para restrição de exclusão (filho nascido em SP) - [Huber e Mellace \(2014\)](#)

Controles	Observações	% compliers	DP	P-valor		
				(1)	(2)	(3)
Amostra completa	28.458	5,2894	0,0013	1,0000	1,0000	1,0000
Sem instrução	15.533	3,6478	-0,0396	1,0000	1,0000	0,5525
Nível fundamental	5.698	5,7076	-0,0465	1,0000	1,0000	0,9175
Nível médio	6.237	6,9459	-0,0288	1,0000	1,0000	1,0000
Nível superior	944	11,0466	0,1420	0,9697	0,9670	0,5470
Pós graduação	46	22,9946	-	0,9948	0,9970	0,9355
Idade entre 18 e 29	11.741	5,5757	0,0844	0,7032	0,936	0,5155
Idade entre 30 e 49	14.188	5,8026	-0,0195	1,0000	1,0000	1,0000
Idade entre 50 e 70	2.529	7,6408	-0,1392	0,9035	0,9675	0,5490
Idade entre 25 e 35 e sem instrução	6.744	4,6615	0,0074	0,9998	1,0000	0,5460
Região de Origem - Nordeste	17.624	2,4381	-0,0599	0,9350	0,9950	0,4935
Região de Origem - Sudeste	5.766	8,1128	0,0940	0,9081	0,9840	0,5085
Região de Origem - Sul	3.994	14,6260	0,1381	0,9859	0,9995	0,5180
Região de Origem - Centro-oeste	772	12,0704	0,0759	0,9934	0,9985	0,5555
Região de Origem - Norte	302	11,7887	0,3166	0,1796	0,2360	0,1170

Fonte: Elaboração própria com base nos dados do Censo Demográfico de 2010.

Nota: DP (distância padronizada das restrições de desigualdade em 5.1. P valores - (1) baseado no teste de [Chen e Szroeter \(2014\)](#). (2) e (3) - Baseados em [Bennett et al. \(2009\)](#) com recentramento completo e parcial de restrições de desigualdade em 5.1, respectivamente. Total de 999 reamostragens *bootstrap*.

Tabela A8 – Equações de seleção e salários - modelos incluindo mulheres

Covariadas	Somente Mulheres		Homens e Mulheres	
	(probit, $S = 1$ )	(Equação de Salários)	(probit, $S = 1$ )	(Equação de Salários)
<i>Não tem filho (categoria omitida)</i>				
Filho de 14 anos nascido em SP	0,1302*** (0,0317)		0,2554*** (0,0197)	
Filho de 14 anos	-0,3366*** (0,0407)	-0,0096 (0,0126)	-0,4330*** (0,0304)	-0,0092 (0,0073)
<i>Branco (categoria omitida)</i>				
Não branco	-0,0181 (0,0284)	-0,0624*** (0,0081)	-0,0205 (0,0170)	-0,0447*** (0,0049)
<i>Homem (categoria omitida)</i>				
Mulher			0,1139*** (0,0201)	-0,1963*** (0,0057)
<i>Sem instrução (categoria omitida)</i>				
Nível fundamental	-0,1393*** (0,0383)	0,0568*** (0,0108)	-0,1751*** (0,0222)	0,0880*** (0,0062)
Nível médio	-0,2913*** (0,0344)	0,1192*** (0,0100)	-0,3025*** (0,0209)	0,1761*** (0,0060)
Nível superior	-0,2986*** (0,0547)	0,7779*** (0,0179)	-0,3343*** (0,0379)	0,7603*** (0,0129)
Pós graduação	-0,4908** (0,2054)	1,0907*** (0,0785)	-0,4237*** (0,1577)	1,0532*** (0,0585)
<i>Cônjuge (categoria omitida)</i>				
Chefe	-0,1321*** (0,0341)	0,0369*** (0,0111)	0,0353 (0,0280)	0,0521*** (0,0074)
<i>Chefe solteiro (categoria omitida)</i>				
Chefe e vive com cônjuge	-0,0631 (0,0452)	-0,0384*** (0,0144)	-0,2434*** (0,0286)	-0,0186** (0,0076)
Chefe e vive com cônjuge nascido em SP	0,9103*** (0,0987)	0,0623*** (0,0210)	0,6065*** (0,0343)	0,0265*** (0,0081)
<i>Emprego sem carteira assinada (categoria omitida)</i>				
Emprego com carteira assinada	0,3041*** (0,0320)	0,0061 (0,0094)	0,4668*** (0,0210)	0,0508*** (0,0065)
Autônomo	-0,3022*** (0,0407)	0,1642*** (0,0145)	-0,1622*** (0,0251)	0,1553*** (0,0090)
Empregador	-0,6875*** (0,1009)	0,3545*** (0,0472)	-0,5304*** (0,0631)	0,3975*** (0,0296)
<i>Não recebe Programa Social (categoria omitida)</i>				
Recebe programa social	-0,5702*** (0,0433)	-0,0910*** (0,0146)	-0,5225*** (0,0388)	-0,0848*** (0,0133)
<i>Setor rural (categoria omitida)</i>				
Setor urbano	0,1262** (0,0542)	0,0984*** (0,0179)	0,1995*** (0,0289)	0,1643*** (0,0098)
<i>Região de origem - Sudeste (categoria omitida)</i>				
Região de origem - Norte	0,3436*** (0,1261)	0,0349 (0,0330)	0,2728*** (0,0816)	0,0219 (0,0216)
Região de origem - Nordeste	0,3943*** (0,0345)	-0,0477*** (0,0101)	0,3250*** (0,0207)	-0,0709*** (0,0062)
Região de Origem - Sul	-0,3920*** (0,0379)	0,0160 (0,0135)	-0,3608*** (0,0239)	0,0140* (0,0085)
Região de Origem- Centro-oeste	-0,1363** (0,0654)	-0,0272 (0,0224)	-0,1452*** (0,0436)	0,0108 (0,0147)
Intercepto	1,2556*** (0,0701)	1,3356*** (0,0225)	0,9902*** (0,0442)	1,4125*** (0,0138)
Idade (spline)	sim	sim		
Total de pessoas no domicílio (spline)	sim	sim		
Observações	16.166	13.818	44.660	38.192
Parâmetro de Dependência Kendall	$\theta = 2.893(2,647,3,19)$ $\tau = 0,5$		$\theta = 1.963(1.865,2.063)$ $\tau = 0,35$	

Fonte: Elaboração própria com base nos dados do censo 2010. Erros-Padrão entre parêntese. O modelo bivariado é do tipo cópula Joe rotacionada a 180°. Níveis de significância: \*10%, \*\* 5%, \*\*\* 1% .



Tabela A9 – Equações de seleção e salários - abordagem paramétrica

Covariadas	Cópula Normal		Cópula Joe 180°	
	(probit, $S = 1$ )	(Equação de Salários)	(probit, $S = 1$ )	(Equação de Salários)
<i>Não tem filho (categoria omitida)</i>				
Filho de 14 anos nascido em SP	0,3298*** (0,0255)		0.262*** (0.0233)	
Filho de 14 anos	-0,4204*** (0,0335)	0,0102 (0,0087)	-0.359*** (0.0319)	-0.00982 (0.00896)
<i>Branco (categoria omitida)</i>				
Não Branco	-0,0302 (0,0214)	-0,0412*** (0,0063)	-0.03867* (0.0203)	-0.0421*** (0.0064)
Idade	-0,0278*** (0,0016)	0,0067*** (0,0004)	-0.0203* (0.0015)	0.0037*** (0.0004)
Idade ao quadrado	0,0002*** (0,0001)	-0,0002*** (0,00003)	2.853E(-05) (7.577E(-05))	-0.0002*** (2.571 E(-05))
<i>Sem instrução (categoria omitida)</i>				
Nível fundamental	-0,1877*** (0,0275)	0,0995*** (0,0079)	-0.1566*** (0.0262)	0.08004*** (0.0082)
Nível médio	-0,3033*** (0,0268)	0,2106*** (0,0079)	-0.2319*** (0.0257)	0.1773*** (0.0082)
Nível superior	-0,3112*** (0,0517)	0,8003*** (0,0187) ( 0.0524)	-0.2298*** (0.0185)	0.7222*** (0.0185)
Pós graduação	-0,2986 (0,2198)	1,1183*** (0,0789)	-0.2433 (0.2484)	1.038*** (0.0783)
<i>Cônjuge (categoria omitida)</i>				
Chefe	0,1580*** (0,0395)	0,0634*** (0,0108)	0.138*** (0.0376)	0.0698*** (0.0111)
<i>Chefe e não vive com cônjuge (categoria omitida)</i>				
Chefe e vive com cônjuge	-0,3393*** (0,0353)	-0,0127 (0,0097)	-0.306*** (0.0337)	-0.0426*** (0.01)
Chefe e vive com cônjuge nascido em SP	0,5360*** (0,0370)	0,0154* (0,0091)	0.492*** (0.0348)	0.0743*** (0.0098)
<i>Emprego sem carteira assinada (categoria omitida)</i>				
Emprego com carteira assinada	0,5844*** (0,0283)	0,0768*** (0,0095)	0.49*** (0.0276)	0.142*** (0.0098)
Autônomo	-0,0584* (0,0328)	0,1749*** (0,0123)	-0.0443 (0.1490)	0.149*** (0.0119)
Empregador	-0,4252*** (0,0789)	0,4456*** (0,0382)	-0.479*** (0.0785)	0.367*** (0.0355)
<i>Não recebe programa social (categoria omitida)</i>				
Recebe programa social	-0,4375*** (0,1081)	-0,1017*** (0,0394)	-0.39*** (0.106)	-1.49*** (0.0387)
Total de Pessoas no Domicílio	0,1061*** (0,0100)	-0,0076*** (0,0025)	0.0954*** (0.00957)	0.0011 (0.00265)
Total de Pessoas no Domicílio ao quadrado	-0,0031*** (0,0010)	0,0006*** (0,0002)	-0.0029** (0.0011)	0.00034 (0.0002)
<i>Setor Rural (categoria omitida)</i>				
Setor Urbano	0,2279*** (0,0348)	0,1955*** (0,0122)	0.226*** (0.0332)	0.213*** (0,012)
<i>Região de origem - Sudeste (categoria omitida)</i>				
Região de Origem - Norte	0,1814* (0,1054)	0,0135 (0,0292)	0.177* (0,1)	0.0356 ( 0.0301)
Região de Origem - Nordeste	0,2887*** (0,0261)	-0,0797*** (0,0080)	0.229*** (0.025)	-0.0472*** ( 0.0082)
Região de Origem - Sul	-0,3665*** (0,0307)	0,0075 (0,0112)	-0.361*** (0.0298)	-0.0353** (0.011)

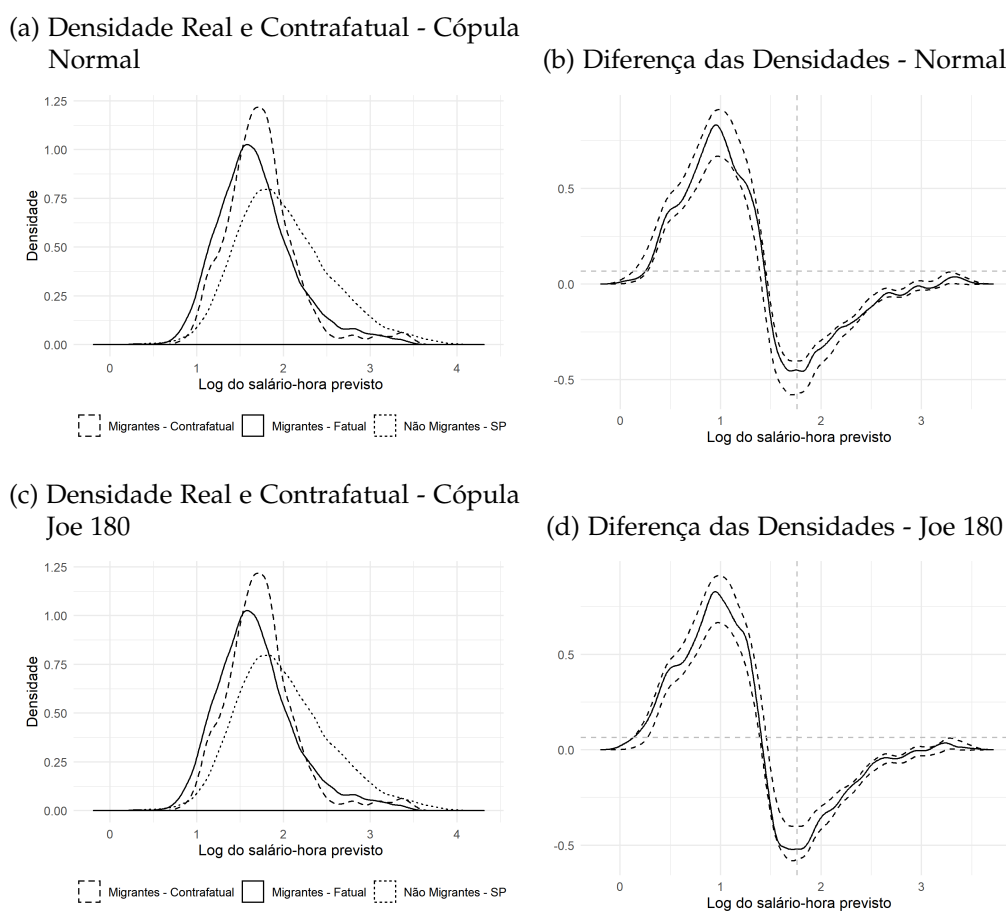
continua na próxima página...

(continuação da página anterior)

Covariadas	Cópula Normal		Cópula Joe 180°	
	(probit, $S = 1$ )	(Equação de Salários)	(probit, $S = 1$ )	(Equação de Salários)
Região de Origem - Cento-oeste	-0,1507*** (0,0585)	0,0361* (0,0200)	-0,136* (0,0562)	0,0169 (0,0197)
Intercepto	1,3924*** (0,0758)	1,1815*** (0,0226)	1,18*** (0,0728)	1,13*** (0,023)
Observações	28.458	24.421	28.458	24.421
Parâmetro de Dependência <i>Kendall</i>	$\theta = 0,01(-0,057,0,08)$ $\tau = 0,006$		$\theta = 2,153(2,011,2,306)$ $\tau = 0,39$	

Fonte: Elaboração própria com base nos dados do Censo Demográfico de 2010. Os modelos (1) e (2) são equivalentes ao modelo de Heckman (1979) estimado por máxima verossimilhança. Erros-Padrão entre parêntese. Níveis de significância: \*10%, \*\* 5%, \*\*\* 1% .

Figura A1 – Densidade real e contrafactual do salário-hora para migrantes e paulistas - abordagem paramétrica



Fonte: Elaboração própria com base nos dados do Censo Demográfico de 2010.

Nota: (a) e (b): as previsões dos salários foram baseadas na cópula Normal com especificação paramétrica, equivalente ao modelo de Heckman (1979) em dois estágio - Tobit2. (c) e (d): previsões de salários baseadas nos parâmetros estimados para o modelo bivariado paramétrico do tipo cópula Joe rotacionada a 180°

Nota 2: Intervalos de confiança representado pelas linhas pontilhadas, supondo um nível de 95% de confiança. A estimação dos erros-padrão foi feita usando a técnica de reamostragem *bootstrap*.

Nota 3: O teste Kolmogorov - Smirnov rejeitou-se a hipótese nula de que as densidades reais e contrafatuais sejam iguais a 1% de significância em ambos os modelos

Tabela A10 – Estimativas das equações de seleção e salários - modelo sem restrição de exclusão (Filho nascido em SP)

Covariadas	Equação de Seleção ( <i>probit</i> , $S = 1$ ) (1)	Equação de Salários (2)
<i>Não tem filho (categoria omitida)</i>		
Filho de 14 anos	-0,2960*** (0,0379)	-0,0001 (0,0093)
<i>Branco (categoria omitida)</i>		
Não Branco	-0,0335 (0,0214)	-0,0408*** (0,0063)
<i>Sem Instrução (categoria omitida)</i>		
Nível Fundamental	-0,1902*** (0,0275)	0,1006*** (0,0079)
Nível Médio	-0,3116*** (0,0268)	0,2098*** (0,0080)
Nível superior	-0,3139*** (0,0516)	0,7989*** (0,0187)
Pós-Graduação	-0,3011 (0,2176)	1,1171*** (0,0788)
<i>Cônjuge (categoria omitida)</i>		
Chefe	0,2605*** (0,0462)	0,0692*** (0,0110)
<i>Chefe Solteiro (categoria omitida)</i>		
Chefe e vive com cônjuge	-0,4645*** (0,0426)	-0,0189* (0,0099)
Chefe e vive com cônjuge nascido em SP	0,6402*** (0,0360)	0,0161* (0,0091)
<i>Emprego sem carteira assinada (categoria omitida)</i>		
Emprego com carteira assinada	0,5806*** (0,0283)	0,0761*** (0,0095)
Autônomo	-0,0399 (0,0327)	0,1749*** (0,0122)
Empregador	-0,3775*** (0,0788)	0,4421*** (0,0382)
<i>Não recebe programa social (categoria omitida)</i>		
Recebe programa social	-0,4321*** (0,1091)	-0,1046*** (0,0393)
<i>Setor rural (categoria omitida)</i>		
Setor urbano	0,2406*** (0,0347)	0,1956*** (0,0122)
<i>Região de origem - Sudeste (categoria omitida)</i>		
Região de origem - Norte	0,1753* (0,1057)	0,0141 (0,0292)
Região de origem - Nordeste	0,2971*** (0,0260)	-0,0802*** (0,0080)
Região de origem - Sul	-0,3554*** (0,0306)	0,0076 (0,0112)
Região de origem - Centro-Oeste	-0,1575*** (0,0583)	0,0353* (0,0200)
Intercepto	0,8527*** (0,0570)	1,3668*** (0,0182)
Idade ( <i>spline</i> )	sim	sim
Total de pessoas no domicílio ( <i>spline</i> )	sim	sim
Observações	28.458	24.421
Parâmetro de Dependência (Intervalo de confiança 95%): $\theta = 2.152(2.02, 2.32)$		
$\tau$ de Kendall : $\tau = 0.38$		

Fonte: Elaboração própria com base nos dados do censo 2010.

Nota 1: Erros-Padrão entre parêntese. : O modelo bivariado é do tipo cópula *Joe* rotacionada a 180°. Níveis de significância: \*10%, \*\* 5%, \*\*\* 1%

Tabela A11 – Regressões salariais de não migrantes paulistas

Covariadas	Modelo principal	Amostra de mulheres	Homens e mulheres
	(1) Fig. 4	(2) Fig. 6 (a)	(3) Fig. 6 (c)
<i>Homem (categoria omitida)</i>			
Mulher	-	-	-0,2243*** (0,0028)
<i>Branco (categoria omitida)</i>			
Não branco	-0,0955*** (0,0032)	-0,0675*** (0,0039)	-0,0860*** (0,0025)
<i>Sem instrução (categoria omitida)</i>			
Nível Fundamental	0,1781*** (0,0043)	0,1086*** (0,0054)	0,1564*** (0,0033)
Nível Médio	0,3550*** (0,0038)	0,2503*** (0,0047)	0,3202*** (0,0029)
Nível superior	0,9085*** (0,0052)	0,8974*** (0,0057)	0,9006*** (0,0038)
Pós-Graduação	1,1929*** (0,0212)	1,2608*** (0,0220)	1,2072*** (0,0157)
<i>Cônjuge (categoria omitida)</i>			
Chefe	-0,0003 (0,0089)	0,0386*** (0,0121)	0,0144** (0,0061)
<i>Chefe Solteiro (categoria omitida)</i>			
Chefe e vive com cônjuge	0,0518*** (0,0080)	-0,0014 (0,0109)	0,0408*** (0,0057)
Chefe e vive com cônjuge nascido em SP	0,0217*** (0,0049)	0,0247** (0,0110)	0,0197*** (0,0045)
<i>Não tem filho (categoria omitida)</i>			
Filho de 14 anos	0,0374*** (0,0059)	-0,0010 (0,0066)	0,0237*** (0,0044)
<i>Emprego sem carteira assinada (categoria omitida)</i>			
Emprego com carteira assinada	0,0930*** (0,0049)	0,0349*** (0,0047)	0,0760*** (0,0033)
Autônomo	0,1859*** (0,0054)	0,2195*** (0,0057)	0,1907*** (0,0038)
Empregador	0,4694*** (0,0091)	0,5141*** (0,0115)	0,4847*** (0,0071)
<i>Não recebe programa social (categoria omitida)</i>			
Recebe programa social	-0,2048*** (0,0296)	-0,1205*** (0,0094)	-0,1098*** (0,0089)
<i>Setor Rural (categoria omitida)</i>			
Setor Urbano	0,2100*** (0,0053)	0,1117*** (0,0074)	0,1881*** (0,0042)
Intercepto	1,3783*** (0,0099)	1,2905*** (0,0144)	1,3920*** (0,0077)
Idade ( <i>spline</i> )	sim	sim	
Total de pessoas no domicílio ( <i>spline</i> )	sim	sim	
Observações	140.166	95.960	237.438
R <sup>2</sup> ajustado	0,2842	0,3066	0,3170

Fonte: Elaboração própria com base nos dados do Censo Demográfico de 2010.

Nota: Os parâmetros das equações salariais 1, 2 e 3 foram estimados a partir de modelos aditivos generalizados com estimativa integrada de suavidade, mais detalhes em Wood e Wood (2015).