

AUTOSSELEÇÃO E DINÂMICA INTERGERACIONAL DE EDUCAÇÃO NO BRASIL

José Luis da Silva Netto Junior (PPGE/UFPB)
Hilton Martins de Britto Ramalho (PPGE/UFPB)

Resumo

O artigo analisa os determinantes da mobilidade intergeracional da educação no Brasil a partir de um modelo estrutural com mudança endógena de regime aplicado aos dados da PNAD de 1996. Dada à variedade de trabalhos na literatura que ignoram a possibilidade de a amostra distribuir-se de forma não aleatória, o presente estudo aborda duas questões centrais. Primeiro, busca-se verificar se a ausência de filhos adultos co-residentes com os pais torna o coeficiente de mobilidade tendencioso. Em segundo lugar, procura-se investigar o grau de inter-relação entre a decisão de permanência do filho no domicílio durante a vida adulta e a expectativa de acumulação de capital humano. Os resultados sugerem a persistência intergeracional da educação é menor para os filhos co-residentes que para aqueles emancipados quando corrigido o viés de seleção amostral. Desse modo, parâmetros tradicionalmente estimados apenas para filhos emancipados e por MQO não podem ser generalizados. Também foi constatada uma relação direta entre a decisão de co-residência com os pais na fase adulta e o grau de instrução acumulado e que a educação dos pais favorece a permanência dos filhos no lar.

Palavras-Chave: *Autosseleção; Capital Humano; Regressão de Mudança Endógena.*

Abstract

The article analyzes the determinants of intergenerational mobility of education in Brazil from an endogenous switching regression applied to data from PNAD 1996. Given the variety of works in the literature ignore the possibility of distributing the sample is not randomly, this study addresses two central questions. First, we seek to verify the absence of co-resident adult children with their parents makes the coefficient of mobility bias. Secondly, it seeks to investigate the degree of interrelationship between the child's decision to stay at home during the adult life expectancy and human capital accumulation. The results suggest the intergenerational persistence of education is lower for the children co-resident for those emancipated when corrected for sample selection bias. Thus, parameters traditionally estimated only for children and emancipated by OLS can not be generalized. It was also found a direct relationship between the decision to co-reside with their parents in adulthood and level of education accumulated and parent education favors the permanence of children at home.

Keywords: *Self-Selection, Human Capital, Endogenous Switching Regression.*

1. Introdução

O processo de acumulação de capital humano tem um importante papel no tocante ao crescimento econômico e distribuição de renda (LUCAS, 1988 e BECKER et al., 1990). Entre os fatores que afetam este processo de acumulação vale destacar a estrutura e as características componentes do ambiente familiar. Nesse contexto, a decisão de investimento em capital humano dos filhos, por exemplo, é em grande parte determinada por escolhas feitas pelos pais. Desse modo, é razoável sugerir que as decisões de natureza familiar têm um grande papel quanto à trajetória educacional dos filhos (BECKER, 1981).

Em paralelo, é importante destacar que a localização espacial também pode influenciar nas decisões quanto à alocação de recursos para investimento em capital humano. Deste modo, em áreas mais dinâmicas economicamente é de se esperar que exista um ambiente mais propício para o investimento em educação que em áreas menos desenvolvidas. Portanto, é razoável assumir a existência de uma interação entre os determinantes familiares, efeito local, e fatores associados à localização espacial, efeito externo, no tocante às decisões de investimento em capital humano (GALOR e ZEIRA, 1996 e GALOR e TSIDDON 1998).

No Brasil, a literatura empírica que trata da dinâmica intergeracional educacional e de renda, explora, em sua maior parte, as informações da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 1996 (FERREIRA E VELOSO, 2003, BEHRMAN et al., 2001, MACHADO e GONZAGA, 2007 e PÈRO E SZERMAN, 2008). O citado suplemento social apresenta informações dos pais dos entrevistados que são pessoas de referência, ou chefes do domicílio, que não moram sob o mesmo teto. Em outras palavras, grande parte dos resultados concernentes à dinâmica intergeracional educacional (FERREIRA E VELOSO, 2003 e BEHRMAN et al., 2001) e de renda (FERREIRA e VELOSO, 2006 PÈRO E SZERMAN, 2008) baseiam-se na inexistência de co-residência entre os pares de pais e filhos. Os filhos, na análise feita pelos citados artigos, são todos emancipados.

Neste contexto, é assumido pelo presente estudo que a não contabilização na análise de mobilidade intergeracional educacional de filhos adultos co-residentes pode gerar vieses. Desta forma, o presente estudo pretende responder duas questões básicas: i) não considerar os filhos adultos que residem com os pais implica em algum viés na análise de mobilidade intergeracional educacional? e ii) de que forma a decisão de permanência no domicílio dos pais está associado com a expectativa de acumulação de capital humano e quais variáveis estão associadas a esta decisão?

Além desta introdução este artigo está dividido em mais cinco seções. Na segunda parte é apresentada uma breve revisão da literatura empírica que trata do tema de estudo. As considerações a respeito do método e base de dados são descritas na terceira seção. A quarta parte apresenta resultados do estudo e a quinta as considerações finais.

2. Revisão da Literatura

Os estudos empíricos que tratam da mobilidade intergeracional baseiam-se na ideia de que existe uma relação entre estoque de capital de capital humano dos pais e filhos. O efeito dos pais sobre a trajetória educacional dos filhos, efeito local, é captado pelas mais variadas medidas de persistência que mensura o quanto os filhos tendem a ser “iguais” aos seus pais. Neste sentido, uma alta persistência sugere uma baixa mobilidade intergeracional enquanto que uma baixa persistência baseia-se numa menor chance dos filhos terem trajetória semelhante dos pais.¹

¹ Uma maior mobilidade é sempre entendida como uma melhora dos indicadores, mas por definição, uma maior mobilidade pode estar associada a piora (FIELDS e OK, 1996). Contudo, é assumido neste artigo que um

Entretanto, a persistência dos efeitos dos pais engloba diversas outras variáveis (lar monoparental ou biparental, sexo, cor, idade etc.) que compõe o chamado efeito familiar. Este conjunto de fatores, por sua vez, interage com aspectos associados ao ambiente externo relacionado com características locais. Deste modo, é de se esperar que em áreas onde exista uma dinâmica econômica mais intensa apresente infraestrutura e incentivos maiores para o investimento em acumulação de capital humano (GALOR e ZEIRA, 1996 e GALOR e TSIDDON 1998).

A literatura empírica que investiga questões relativas à mobilidade intergeracional educacional data, em sua grande parte, do final da década de 1990 e início dos anos 2000. No tocante aos determinantes familiares associados à escolaridade merecem destaque os estudos de Barros e Lam (1993) e Barros et al. (2001). No primeiro é feita uma análise da influência dos pais quanto à trajetória dos filhos com idade de 14 anos, residentes em áreas urbanas, nas regiões Sudeste e Nordeste usando os dados da PNAD de 1982. O último trata da mesma temática só que, desta vez, ampliando a faixa etária para 11 a 25 anos, tendo como base a PNAD de 1996. Contudo, os dois estudos ressaltam a importância das características familiares para o desempenho educacional dos filhos e assimetrias quanto a este efeito do ponto de vista espacial.

O estudo de Berhman et al. (2001) foi um dos primeiros a abordar os impactos das reformas econômicas ocorridas na América Latina sobre os indicadores de pobreza e desigualdade de renda tendo como um dos objetivos estimar o grau de mobilidade intergeracional educacional. Em linhas gerais, em Berhman et al. (2001), tendo como base os dados do suplemento social da PNAD de 1996, são mostradas evidências que sugerem uma elevada persistência educacional no Brasil, ou seja, baixa mobilidade intergeracional educacional mesmo se comparado aos países latino americanos.

No artigo de Ferreira e Veloso (2003), tendo como base também o suplemento social do censo de 1996, observou-se uma alta persistência intergeracional educacional no Brasil se comparado aos demais países latino americanos. Em paralelo, observou-se uma assimetria na relação escolaridade do pai e filho de acordo o nível de escolaridade, região geográfica e características familiares. Vale salientar que o estudo de Berhman et al. (2001) e o de Ferreira e Veloso (2003) não consideraram os filhos co-residentes, ou seja, aqueles que ainda moram com os pais na idade adulta.

Considerando os estudos que assumem apenas os pais e filhos co-residentes merece destaque os artigos Figueiredo et al. (2007) e Netto Jr. et al. (2008). O primeiro utiliza informações das PNAD de 1987 a 2005 e conclui que ao longo deste período houve um aumento da mobilidade intergeracional educacional associada a uma redução da persistência do pai analfabeto. O último estudo tendo como base dados os microdados do Censo Demográfico do IBGE de 2000 ressalta a grande heterogeneidade da dinâmica educacional inter-regional e como ela se diferencia considerando indivíduos migrantes e não migrantes.

Na literatura empírica internacional, que trata da mobilidade intergeracional educacional e de rendimentos, grande parte dos estudos tem como base de dados informações longitudinais de famílias (BORJAS, 1992, MULLIGAN, 1997 e ATKINSON, 1981). Em geral, a preocupação básica destas pesquisas reside na determinação da persistência intergeracional. As principais evidências sugerem que existe uma relação forte entre o nível educacional dos pais e de seus filhos.

Quanto aos determinantes da mobilidade educacional merecem destaque os estudos de Ermisch e Francesconi (2000) e Ermisch e Nicoletti (2006) para Grã-Bretanha. Essas pesquisas enfatizam a importância da escolaridade dos pais e da dotação de riqueza para a

aumento da mobilidade intergeracional educacional está associada a um aumento da média de anos de estudo e redução da desigualdade educacional.

mobilidade intergeracional educacional. De acordos com os resultados de Ermisch e Francesconi (2000), há baixa mobilidade intergeracional em lares monoparentais.

O presente estudo se insere nesta literatura na medida em que discute aspectos associados à mobilidade Intergeracional educacional debatidos nos textos de Berhman *et al.* (2001) e Ferreira e Veloso (2003). Neste sentido discute as consequências de possíveis vieses relativos à inclusão ou não de filhos não emancipados, ou co-residentes, no debate acerca da mobilidade além de fazer considerações a respeito dos possíveis determinantes da decisão de se emancipar ou não.

3. Estratégia Empírica

3.1. O Modelo Empírico

Várias pesquisas internacionais adotaram um modelo de regressão linear simples para captar, a partir de uma medida sintética, o efeito da transmissão educacional dos pais para os filhos (Borjas (1992), Couch e Dunn (1997), Berhman et al. (2001) e Ferreira e Veloso (2003)). Formalmente, o modelo pode ser descrito pela equação a seguir:

$$Y_i^c = \alpha + \beta Y_i^p + \theta X_i + \varepsilon_i \quad (1)$$

Onde: Y_i^c é o nível de instrução do filho, Y_i^p o nível de educação do pai, X_i é um vetor de características socioeconômicas, α e β coeficientes, θ é um vetor de parâmetros e ε_i um termo randômico normalmente distribuído e com variância constante.

Nesse modelo, o coeficiente β tem um papel central, ou seja, fornece uma medida do grau de persistência intergeracional de educação. Desse modo, quanto maior o valor do referido coeficiente, menor deve ser o grau de mobilidade intergeracional. Conforme destacado por Ferreira e Veloso (2003), a medida $1 - \beta$ fornece o grau de mobilidade intergeracional de educação, ou seja, quanto menor seu valor, maior será a mobilidade.

A maioria dos estudos internacionais e nacionais sobre a temática em destaque ignorou a possibilidade de a amostra distribuir-se de forma não aleatória, justamente, por não considerar os filhos adultos que permaneceram no domicílio sob a tutela dos pais. A decisão de permanência do filho no domicílio pode ser condicionada, entre outros fatores, ao grau de instrução dos pais. Nessas circunstâncias, a estimativa tradicional da equação (1), por Mínimos Quadrados Ordinários (MQO), pode produzir parâmetros tendenciosos na presença de uma amostra autosselecionada em atributos não observados (HECKMAN, 1979). Portanto, a depender do viés de seleção amostral, a medida de persistência estimada β pode não representar o verdadeiro parâmetro da população e mascarar comparações internacionais.

Diante desse problema, optou-se pela utilização de um modelo de seleção com mudança endógena (*endogenous switching regression*), procedimento que permite averiguar se o indivíduo decide permanecer ou sair do domicílio chefiado pelo pai de acordo com as vantagens comparativas em termos de investimento em educação.

Considere uma equação de utilidade aleatória U_i que determina a seleção não aleatória dos filhos na amostra, isto é, a equação de decisão entre morar ou não com os pais na fase adulta:

$$U_i = \delta(Y_{1i}^c - Y_{2i}^c) + \alpha Z_i + v_i \quad (2)$$

Onde: Z_i é um vetor de variáveis relacionadas aos custos de moradia; Y_{1i}^c é a escolaridade esperada do filho que decide residir com seus pais durante a vida adulta; Y_{2i}^c o grau de instrução do filho que pretende deixar de viver junto com seus pais quando adulto; δ e α (vetor) representam parâmetros estruturais e v_i um termo de erro probabilístico associado aos fatores não observados e relacionados aos custos de moradia.

O nível de educação dos filhos é apenas observado de acordo com a condição domiciliar escolhida (regime), contudo, condicionado à escolaridade do pai e aos atributos socioeconômicos. Portanto, há duas equações de transmissão intergeracional de educação, respectivamente, para filhos co-residentes (ou filhos dependentes) e não co-residentes (ou filhos emancipados):

$$Y_{1i}^c = \phi_1 + \beta_1 Y_{1i}^p + \theta_1 X_{1i} + \varepsilon_{1i} \quad (3)$$

$$Y_{2i}^c = \phi_2 + \beta_2 Y_{2i}^p + \theta_2 X_{2i} + \varepsilon_{2i} \quad (4)$$

Onde: Y_{1i}^p e Y_{2i}^p representam o grau de instrução do pai em cada regime, X_{1i} e X_{2i} são vetores de atributos pessoais e de localização; ϕ_1 e ϕ_2 coeficientes (interceptos), β_1 e β_2 coeficientes de persistência intergeracional; θ_1 e θ_2 vetores de parâmetros e ε_{1i} e ε_{2i} são termos de erros randômicos.

A equação reduzida do modelo por ser obtida a partir da substituição dessas últimas equações em (2):

$$U_i^* = \delta(\phi_1 + \beta_1 Y_{1i}^p + \theta_1 X_{1i} + \varepsilon_{1i} - \phi_2 - \beta_2 Y_{2i}^p - \theta_2 X_{2i} - \varepsilon_{2i}) + \alpha Z_i + v_i$$

$$U_i^* = (\delta\phi_1 - \delta\phi_2 + \delta\beta_1 Y_{1i}^p - \delta\beta_2 Y_{2i}^p + \delta\theta_1 X_{1i} - \delta\theta_2 X_{2i} + \alpha Z_i) + (\delta\varepsilon_{1i} - \delta\varepsilon_{2i} + v_i) = V_i + \epsilon_i \quad (5)$$

Onde: U_i^* é uma variável latente e contínua (não observada) que mede o benefício por permanecer no domicílio; $V_i \equiv (\delta\phi_1 - \delta\phi_2 + \delta\beta_1 Y_{1i}^p - \delta\beta_2 Y_{2i}^p + \delta\theta_1 X_{1i} - \delta\theta_2 X_{2i} + \alpha Z_i)$ é uma combinação linear dos atributos determinantes dos custos de moradia com os pais e da educação; $\epsilon_i \equiv (\delta\varepsilon_{1i} - \delta\varepsilon_{2i} + v_i)$ captura a combinação dos termos randômicos de (3) e (4).

Portanto, a decisão de permanência do filho no domicílio dos pais dependerá do valor do benefício esperado por essa escolha. Assim, se $U_i^* \geq 0$ o filho deve continuar morando com seus pais na fase adulta, caso contrário $U_i^* < 0$, deverá deixá-los quando adulto.

Assumindo que ϵ_i é normalmente distribuído, Maddala (1983) mostra que é possível estimar a equação (5) e as probabilidades individuais de permanência no domicílio a partir de um *probit* univariado $prob(d = 1) = \Lambda(V_i + \epsilon_i)$ por Máxima Verossimilhança (MV), onde Λ é a função de densidade acumulada da distribuição normal padrão; d é uma variável indicadora (binária) que assume o valor 1 se o filho mora com os pais, e 0 caso contrário.

É importante ressaltar que, devido à seleção amostral ser supostamente não aleatória, os termos de erro de (3) e (4) são correlacionados com o componente aleatório de (5), de modo que estimativas isoladas das duas primeiras equações produziriam coeficientes tendenciosos (CAMERON e TRIVEDI, 2005)². A seguir, discutem-se os métodos mais usuais para a correção do viés de seleção amostral.

3.2. Correção do Viés de Seleção Amostral

Lee (1978) e Heckman (1979) desenvolveram uma técnica de correção para viés de seleção amostral baseada em propriedades da distribuição normal que é bastante usada na literatura, conhecida como *heckit*. A aplicação desse método ao modelo supracitado parte da hipótese que os termos de randômicos de (3), (4) e (5) seguem uma distribuição normal conjunta, com matriz de covariância dada por:

² Note-se que o termo aleatório de (5) é ao menos uma combinação linear dos termos de (3) e (4).

$$\text{cov}(\epsilon, \epsilon_{1i}, \epsilon_{2i}) = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_1 \rho_{1\epsilon} & \sigma_2 \rho_{2\epsilon} \\ \sigma_1 \rho_{1\epsilon} & \sigma_1^2 & . \\ \sigma_2 \rho_{2\epsilon} & . & \sigma_2^2 \end{pmatrix}$$

Onde: $\rho_{1\epsilon} \in [-1,1]$ e $\rho_{2\epsilon} \in [-1,1]$ são, respectivamente, os coeficiente de correlação entre os termos aleatórios de (5) e (3) e (5) e (4), σ_1 é o desvio padrão do erro probabilístico de (3), σ_2 é o desvio padrão do erro estocástico de (4) e σ_ϵ^2 , σ_1^2 e σ_2^2 , sequencialmente, as variâncias dos erros randômicos de (5), (3) e (4).

Destarte, é possível mostrar que a escolaridade esperada do filho, condicionada a sua situação domiciliar, é dada por:

$$E(Y_{1i}^c | d = 1) = \hat{\phi}_1 + \hat{\beta}_1 Y_{1i}^p + \hat{\theta}_1 X_{1i} + \hat{\sigma}_1 \hat{\rho}_{1\epsilon} \frac{f(\hat{V}_i)}{F(\hat{V}_i)} \quad (6)$$

$$E(Y_{2i}^c | d = 0) = \hat{\phi}_2 + \hat{\beta}_2 Y_{2i}^p + \hat{\theta}_2 X_{2i} - \hat{\sigma}_2 \hat{\rho}_{2\epsilon} \frac{f(\hat{V}_i)}{1 - F(\hat{V}_i)} \quad (7)$$

Onde: $\hat{\phi}_1$, $\hat{\beta}_1$, $\hat{\phi}_2$, $(\hat{\theta}_1, \hat{\theta}_2)$ e $\hat{\beta}_2$ são coeficientes (vetores de parâmetros) corrigidos para viés de seleção amostral; $\hat{\sigma}_1 \hat{\rho}_{1\epsilon}$ e $\hat{\sigma}_2 \hat{\rho}_{2\epsilon}$ estimativas das covariâncias entre os erros de (3) e (5) e (4) e (5), respectivamente; \hat{V}_i é a predição linear da equação (5); f é a função de densidade normal e F a função de densidade acumulada. Em particular, os termos $\frac{f(\hat{V}_i)}{F(\hat{V}_i)}$ e $\frac{f(\hat{V}_i)}{1-F(\hat{V}_i)}$ são medidas de probabilidade de seleção amostral (termos de correção), conhecidas como razões inversas de *Mills*.

Portanto, o método *heckit* consiste em estimativas de dois estágios. Primeiro, faz-se a regressão de (5) por um *probit* contemplando toda amostra de filhos e usam-se as predições lineares para computar as razões inversas de *Mills* (termos de correção). Finalmente, no segundo estágio, estimam-se as equações (3) e (4) por MQO em amostras separadas por condição do filho do domicílio, imputando-se os termos de correção como regressores adicionais (vide equações (6) e (7)).

Todavia, Hall (2002) e Loskshin e Sajaia (2004) mostram que sob as mesmas condições de distribuição normal, a estimativa conjunta de (3), (4) e (5) por Máxima Verossimilhança com Informação Completa (FIML) é preferível, uma vez que produz desvios padrão eficientes em grandes amostras. Logo, priorizou-se esse método de estimação apesar da sua maior exigência computacional.

Por fim, a estimativa dos parâmetros estruturais do modelo empírico pode ser feita em outro estágio, ou seja, tomando-se as predições não condicionais de (3) e (4) para toda amostra e computando-as na equação de utilidade randômica (2). Esta última, por seu turno, deve ser estimada por um *probit* com controle para os custos associados à permanência do filho na residência dos pais³. Dessa forma, caso se encontre $\delta > 0$, por exemplo, a decisão de os filhos residirem com seus pais na fase adulta seria consistente com o retorno esperado em acumulação de capital humano.

4. Base de Dados e Tratamentos

³ Para assegurar a identificação dos parâmetros estruturais da equação (2), há variáveis presentes no vetor Z que não são incluídas nos vetores X_1 e X_2 (restrição por exclusão). A escolha das variáveis a serem excluídas é arbitrária, contudo, vale priorizar a exclusão daquelas que se teoricamente ou estatisticamente se correlacionam apenas com a decisão de permanência no domicílio e não explicam a acumulação de capital humano. Vide mais detalhes na seção de tratamento dos dados.

Os dados utilizados nesse trabalho são oriundos da Pesquisa Nacional por Amostra de Domicílio (PNAD) de 1996. Esta é a mais recente pesquisa a oferecer dados suplementares acerca dos pais das pessoas entrevistadas na condição de chefia familiar. Em especial, o interesse recai no cruzamento das informações sobre o nível de instrução dos pais e dos filhos, assim como, em alguns atributos socioeconômicos dos entrevistados, a saber: gênero, raça, localização, atividade econômica e condição no domicílio.

Com o intuito de se comparar alguns resultados empíricos, sobretudo, no tocante às evidências elencadas por Ferreira e Veloso (2003) e outros estudos internacionais, a amostra foi selecionada de acordo com certos critérios empregados pelos últimos autores. Primeiro, foram consideradas apenas as pessoas do sexo masculino que, na data da pesquisa, se achavam na faixa etária dos 25 aos 64 anos de idade⁴. Desse total, excluíram-se aqueles que não informaram sua própria escolaridade e a do seu pai.

A partir desse recorte inicial, os filhos foram separados em dois grupos: (i) *filhos emancipados* - aqueles que se declararam responsáveis pelo domicílio ou cônjuge, totalizando uma subamostra idêntica a usada por Ferreira e Veloso (2003), isto é, 43.772 indivíduos e (ii) *filhos dependentes ou co-residentes* - aqueles que ainda se encontravam residindo com seus pais no domicílio, ou seja, 8.662 pessoas⁵. Portanto, a amostra final foi de 52.394 filhos, 16% dependentes e 84% emancipados.

Vale ressaltar que os valores da escolaridade dos pais foram compatibilizados de acordo com o mesmo critério empregado no estudo de Ferreira e Veloso (2003). A tabela A.1, no apêndice, fornece uma descrição detalhada de todas as variáveis selecionadas.

A tabela 1, a seguir, apresenta as estatísticas descritivas das amostras selecionadas segundo a condição domiciliar dos filhos. No caso das variáveis dicotômicas, são apresentadas apenas as categorias não omitidas, sendo que nesse caso a média deve ser interpretada como a proporção em relação à categoria omitida.

Tabela 1: Estatísticas descritivas da amostra

Variáveis	Filhos dependentes (co-residentes)				Filhos emancipados (não co-residentes)			
	Média	Desvio Padrão	Mínimo	Máximo	Média	Desvio Padrão	Mínimo	Máximo
Educação Filho	6,77	4,57	0	16	6,03	4,79	0	16
Branco ^d	0,51	0,50	0	1	0,57	0,50	0	1
Idade	31,64	6,68	25	63	41,82	10,40	25	64
Idade2	1.045,91	492,02	625	3.969	1.857,04	911,56	625	4.096
Urbana ^d	0,82	0,38	0	1	0,81	0,39	0	1
Educação Pai	3,46	3,98	0	16	2,64	3,53	0	16
Econ. Ativo ^d	0,85	0,36	0	1	0,92	0,27	0	1
Moradores	5,20	2,42	2	21	4,22	1,87	1	21
Norte ^d	0,06	0,24	0	1	0,06	0,24	0	1
Nordeste ^d	0,31	0,46	0	1	0,27	0,44	0	1
Sul ^d	0,15	0,36	0	1	0,20	0,40	0	1
Centro-Oeste ^d	0,08	0,27	0	1	0,11	0,31	0	1
Observações	8.622				43.772			

Fonte: Elaboração própria a partir dos dados da PNAD de 1996.

⁴ Há duas justificativas importantes para essa filtragem. Primeiro, a seleção aqui adotada é semelhante à empregada em vários estudos internacionais sobre o tema. Segundo, o recorte pela referida faixa etária procura captar pessoas que já alcançaram determinado ciclo no processo de acumulação de capital humano.

⁵ Os atributos dos filhos residentes no domicílio foram relacionados com a escolaridade dos pais a partir da identificação do código do domicílio e da unidade da federação.

Notas: ^d Variável binária. Para variáveis binárias, a média refere-se à proporção da categoria não omitida.

Os dados revelam diferenças importantes nas características pessoais dos dois grupos em foco. Cabe notar que as médias de instrução do filho co-residente e do seu pai superam aquelas registradas para os filhos emancipados. Outro atributo bastante distintivo é a idade. O filho emancipado típico tem, em média, 42 anos, isto é, 10 anos a mais que o filho dependente. Os filhos emancipados também se acham em um domicílio com menor média de pessoas, sendo que 92% destes trabalhavam ou procuravam emprego (economicamente ativos) contra 85% dos filhos dependentes. Já distribuição dos dois grupos de filhos entre os meios rural e urbano e grandes regiões parece ser mais homogênea. Destaca-se apenas a maior proporção relativa de filhos dependentes nas regiões Nordeste (31%) e Sudeste (40% - categoria omitida) e de filhos emancipados na região Sul (20%).

Cabe ressaltar que as variáveis *Econ.Ativo* e *Moradores* apenas foram incluídas na equação reduzida do modelo estrutural (5) e não pertencem ao conjunto de argumentos das equações de acumulação de capital humano (3) e (4)⁶. Essas exclusões formam uma restrição necessária para a identificação dos parâmetros estruturais da equação (2).

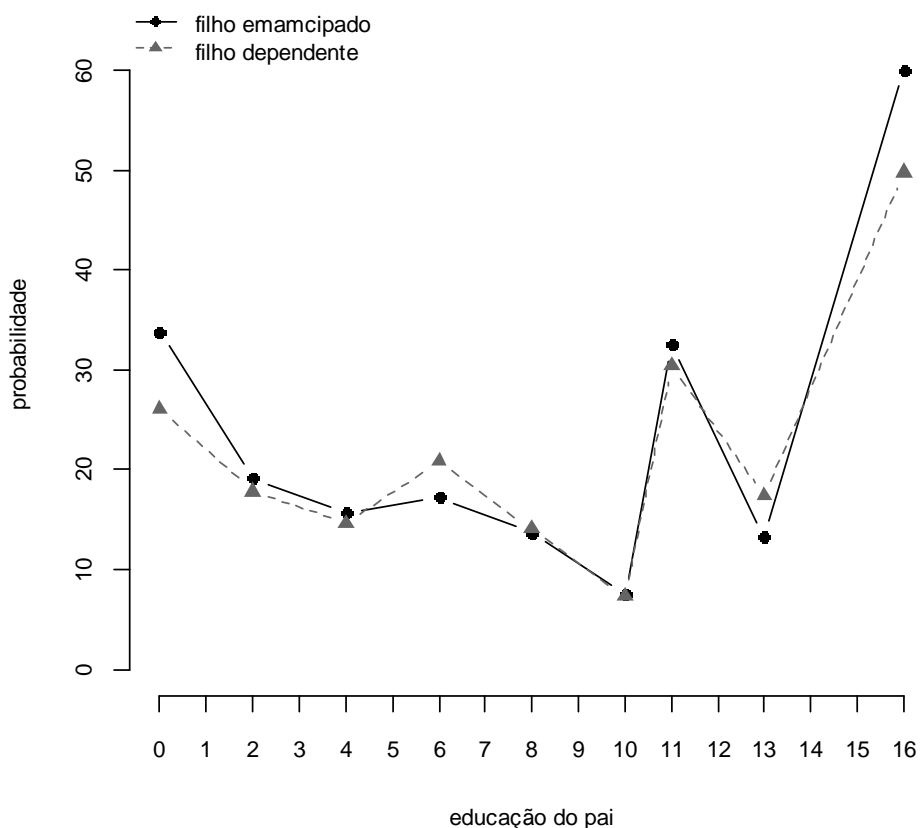
5. Evidências sobre a Mobilidade Intergeracional de Educação no Brasil

5.1. Resultados a partir da análise multivariada com matrizes de transição

A figura 1 registra as probabilidades de persistência educacional dos filhos considerando diferentes faixas de escolaridade do pai e a co-residência com os pais na fase adulta, ou seja, revela as probabilidades condicionadas do filho permanecer com o mesmo nível de instrução do pai, conforme o cálculo das matrizes de transição *markoviana*.

Figura 1: Brasil - Probabilidades de persistência educacional segundo a instrução do pai e por condição do filho no domicílio

⁶ Verificou-se que essas variáveis apresentam baixa correlação com a escolaridade dos filhos e, em geral, não se mostraram estatisticamente significativas quanto incluídas em regressões feitas para (3) e (4).



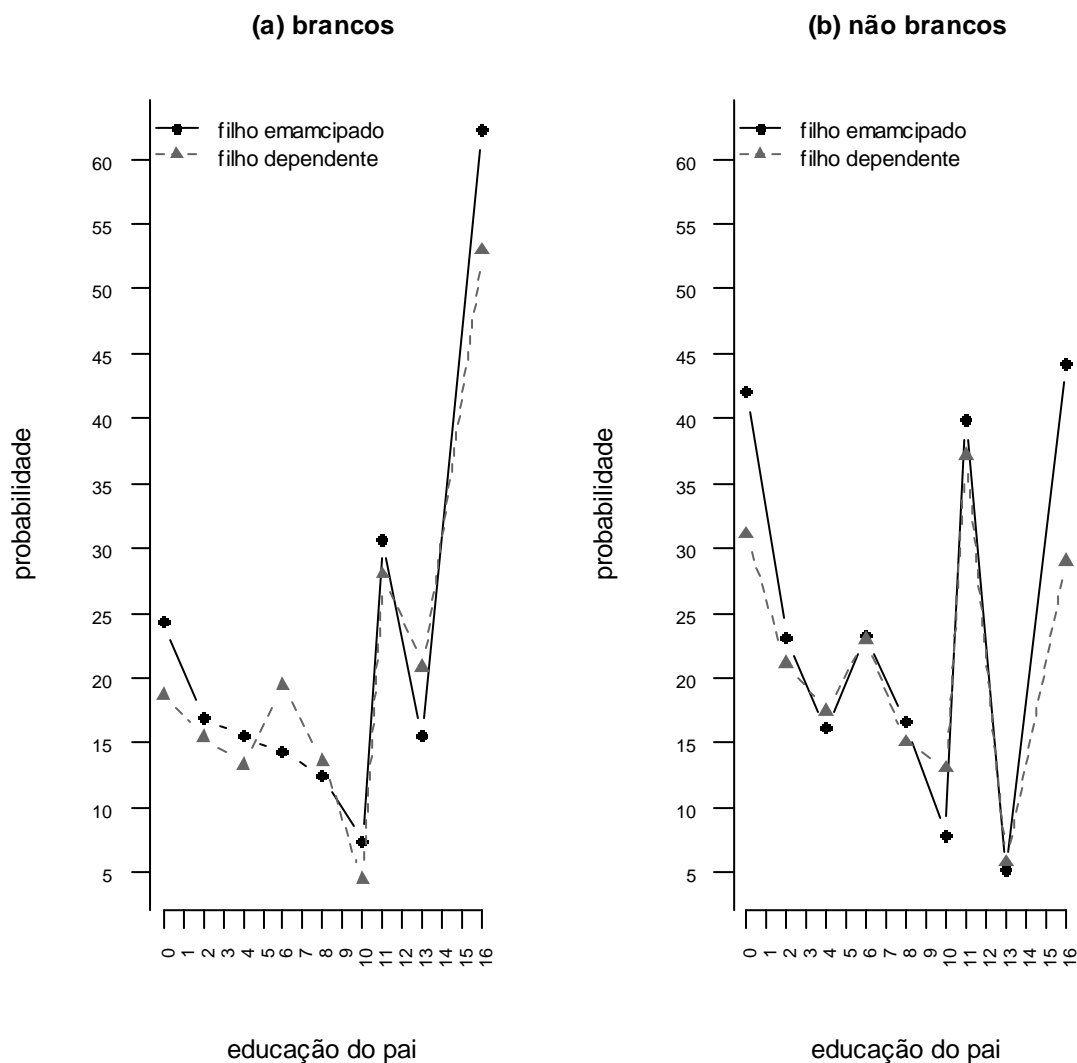
Fonte: Elaboração própria a partir dos dados da PNAD de 1996.

É importante observar que o estoque de capital humano dos pais parece gerar diferenças importantes na mobilidade educacional dos filhos no Brasil, sobretudo, nos extremos da distribuição da instrução, onde a condição de co-residência produz maior diferença. Note-se, por exemplo, que um filho de pai analfabeto tem 33% de probabilidade de permanecer analfabeto, caso não resida com o mesmo na fase adulta (emancipado), e 25%, caso contrário. Por outro lado, um filho de pai com formação superior, registra 60% de chance de acumular esse mesmo nível de instrução em caso de emancipação e 50% se co-residir com os pais quando adulto.

Ainda é possível observar dois outros picos importantes de probabilidade de persistência de educação, ou seja, aqueles condicionados aos pais com 6 anos de instrução ou com 11 anos. Tais resultados, possivelmente, estão correlacionados ao ciclo de conclusão do ensino fundamental e médio, onde muitas pessoas optam por deixar a escola.

A condição socioeconômica das famílias pode ser determinada por uma série de fatores pessoais, familiares e contextuais. É fato muito conhecido na literatura, que a condição social no Brasil é historicamente associada à raça do indivíduo e constitui-se uma importante restrição para o progresso econômico, sobretudo, para as pessoas não brancas. Na figura 2, a seguir, apresentam-se as probabilidades de persistência educacional dos filhos condicionadas à cor da pele, à co-residência com os pais na vida adulta e à própria educação dos pais.

Figura 2: Brasil - Probabilidades de persistência educacional segundo a instrução do pai e por condição do filho no domicílio e raça

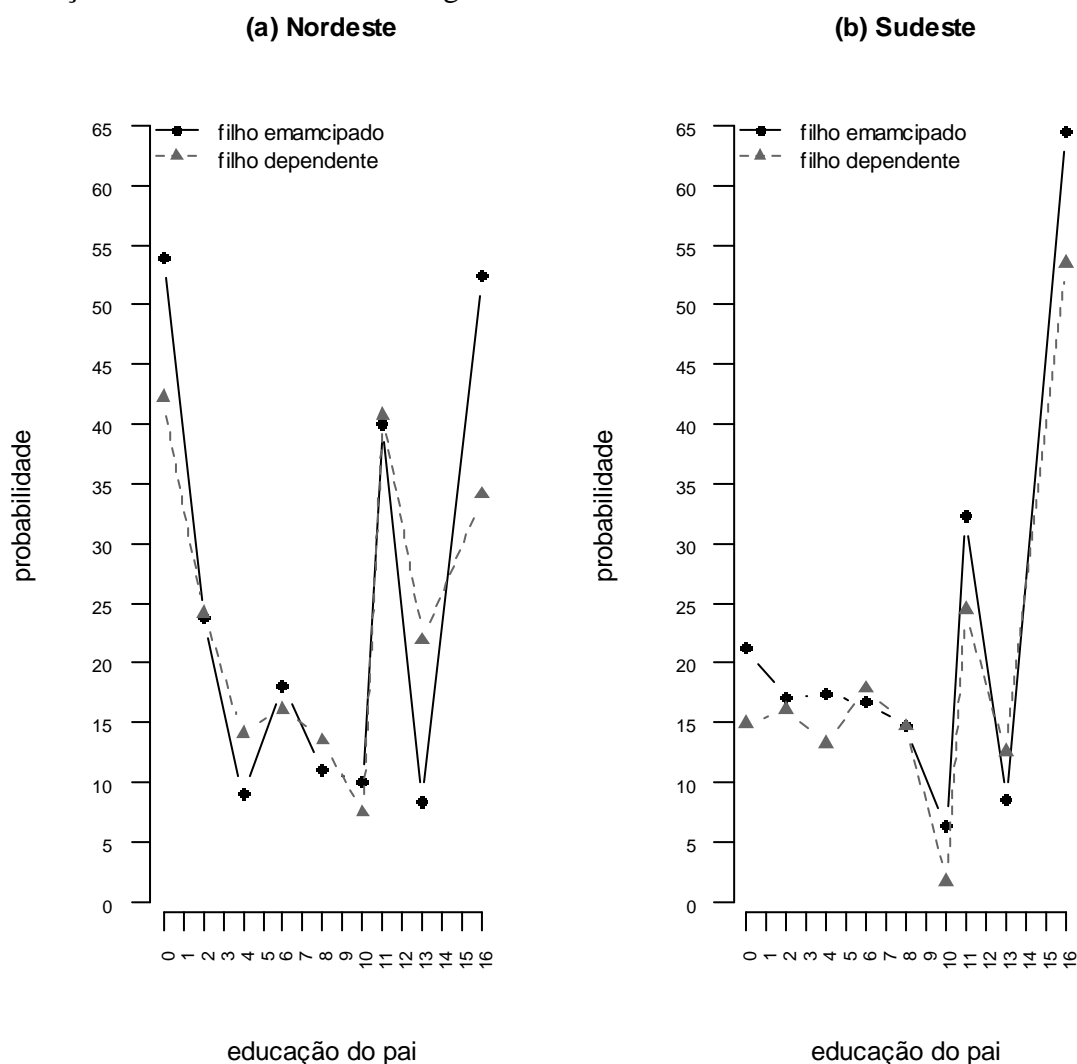


Fonte: Elaboração própria a partir dos dados da PNAD de 1996.

O efeito da raça parece superar consideravelmente a condição do filho no domicílio quando adulto. Enquanto os filhos de cor branca apresentam probabilidades de persistência educacional bastante próximas daquelas observadas para a população em geral, aqueles de cor não branca são relativamente menos favorecidos. Por exemplo, um filho não branco com pai analfabeto registra probabilidade de permanecer com a mesma escolaridade do pai bem superior aquela imputada ao filho típico de cor branca, a despeito da co-residência com os pais quando adulto. No outro extremo, nota-se que um filho não branco cujo pai tem formação superior, apresenta probabilidade de persistência bem inferior àquela observada para um filho de cor branca na mesma condição. Tal diferença ainda é mais acentuada e menos favorável para o filho que co-habita com os pais na fase adulta.

O contexto regional certamente pode moldar o conjunto de oportunidades que os indivíduos se defrontam ao longo da vida, criando, portanto, importantes desigualdades de oportunidades. A figura 3, abaixo, registra as probabilidades de persistência educacional dos filhos considerando duas regiões brasileiras bastante distintas na ótica institucional e econômica: Nordeste e Sudeste. Enquanto a primeira é conhecida pela elevada incidência de pobreza, baixa educação e desigualdade de renda, a segunda concentra os maiores parques industriais e instituições financeiras, respondendo por mais de 50% da riqueza produzida no Brasil.

Figura 3: Brasil - Probabilidades de persistência educacional segundo a instrução do pai e por condição do filho no domicílio e regiões selecionadas

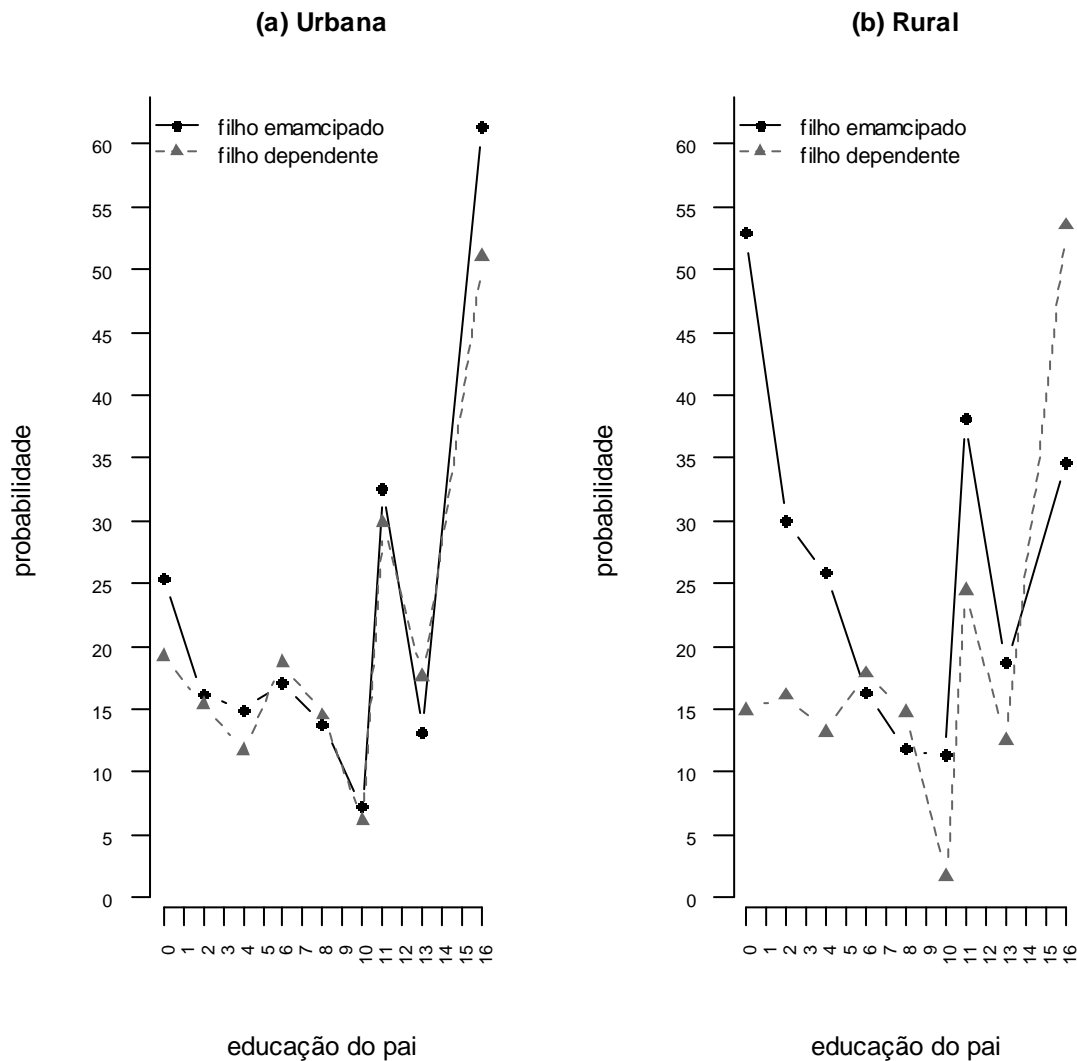


Fonte: Elaboração própria a partir dos dados da PNAD de 1996.

Os resultados revelam que impacto do contexto regional sobre a probabilidade do filho terminar com a mesma escolaridade do seu pai não pode ser desprezado. Os filhos de pai analfabeto no Nordeste são extremamente desfavorecidos aqueles residentes na região Sudeste, sobretudo, se não co-habitarem com o mesmo na vida adulta. Quando se observa a probabilidade de persistência educacional de um filho de pai com nível superior residente no Sudeste, também é percebida a desvantagem de um filho nordestino nessa mesma condição, especialmente, se ele não morar com o pai quando adulto. No Nordeste, apenas os filhos cujo pai registra nível médio de escolaridade apresentam maior probabilidade de persistência quando comparados a um grupo equivalente residindo no Sudeste.

As diferenças nas chances de obter a mesma escolaridade do pai parecem ser ainda mais acirradas quando se considera a dicotomia entre os meios rural e urbano no Brasil. A figura 4, a seguir, revela as probabilidades de persistência educacional dos filhos por setor censitário.

Figura 4: Brasil- Probabilidades de persistência educacional segundo a instrução do pai e por condição do filho no domicílio e setor censitário



Fonte: Elaboração própria a partir dos dados da PNAD de 1996.

Nota: Resultados não incluem a região Norte, onde não foram entrevistados os residentes do meio rural.

É possível constatar que um filho de pai sem instrução que reside na zona rural tem mais de 50% de chance de ser analfabeto quando adulto, especialmente, se for uma pessoa emancipada, contra cerca de 25% de chance de um filho residente no meio urbano. Por outro lado, um filho de pai com curso superior completo residente em área urbana tem 61% de probabilidade de terminar com a mesma escolaridade do seu genitor, caso seja um adulto emancipado, enquanto, um filho sob essas condições no meio rural registra apenas 34% de chance. Contudo, há uma ligeira vantagem relativa do filho cujo pai tem nível médio de escolaridade e reside no meio rural. Nesse caso, sua probabilidade de persistência educacional superar aquela verificada para um filho típico do meio urbano que possui um pai com a mesma instrução reportada.

Em linhas gerais os resultados sugerem que em áreas mais dinâmicas (Sudeste e zonas urbanas) o efeito familiar é menor se comparado àquelas áreas onde os pais registram menor grau de instrução. Contudo, vale observar a assimetria deste efeito considerando ou não a condição de co-residência do filho. Desse modo, as evidências acima reportadas se coadunam com os modelos teóricos propostos por Galor e Zeira (1993) e Galor e Tsiddon (1997) que ressaltam a interação entre o ambiente familiar e o ambiente externo na dinâmica de persistência educacional.

5.2. Resultados empíricos a partir de um modelo estrutural

Nessa subseção, são apresentados os resultados das estimativas paramétricas envolvendo um modelo estrutural de determinação conjunta da acumulação de capital humano da condição domiciliar do filho na vida adulta. Desse modo, na tabela 2, abaixo, as colunas (3), (4) e (5) reportam os resultados da estimação do modelo estrutural na forma reduzida (equações 3, 4 e 5), por MV, enquanto, as colunas (1) e (2), registram, para fins comparativos, regressões por MQO para as equações (3) e (4) supondo distribuição aleatória da amostra.

Os resultados revelam que o coeficiente de persistência intergeracional é relativamente menor para os filhos que residem com os pais na fase adulta (filhos dependentes). Após a correção do viés de seleção amostral, percebe-se um ligeiro aumento dos coeficientes de persistência para os dois grupos de filhos. Tal resultado reforça a hipótese de que a acumulação de capital humano dos filhos está correlacionada com a decisão de viver ou não com os pais.

Nesse sentido, outra evidência relevante é a correlação direta entre a escolaridade do pai e a probabilidade de o filho co-residir com o mesmo na vida adulta, conforme pode ser observado na equação de seleção (coluna (5)). Também se constata que o aumento da idade do filho reduz a probabilidade de permanência com seus pais. Os adultos de cor branca e/ou residentes na zona tem menor chance de permanecer com os pais, quando comparados aos não brancos e moradores do meio rural, categorias omitidas.

A condição de atividade econômica do filho reduz a chance de residência com os pais, comparado ao filho não economicamente ativo, ao passo que o maior número de moradores favorece a co-residência. Residir nas regiões Norte, Nordeste, Sul ou Centro-Oeste aumenta a probabilidade do filho deixar de viver com os pais, quanto se compara a um residente típico da região Sudeste (categoria base)⁷.

Os sinais positivos dos coeficientes de correlação entre os termos aleatórios das equações (3) e (5) e (4) e (5), respectivamente, ρ_1 e ρ_2 sugerem informações sobre a autoseleção dos filhos em atributos não observados. Ou seja, enquanto um filho dependente dos pais acumula mais capital humano que um indivíduo aleatoriamente tomado na amostra, um filho típico emancipado obtém menor instrução. Logo, os resultados sugerem que os filhos co-residentes (emancipados) são positivamente (negativamente) selecionados em atributos não observados.

Algumas características observadas contribuem de forma diferenciada para a instrução dos filhos a depender de sua condição no domicílio. Este é caso do gênero e da idade, uma que um adulto de cor branca tem, em média, mais anos de estudo que uma pessoa não branca, sobretudo, se for emancipado. O aumento da idade favorece a acumulação de capital humano do filho que não reside com os pais e registra efeito oposto para o filho dependente.

Os filhos residentes no meio urbano são, em média, mais instruídos que os residentes na zona rural (categoria omitida), enquanto aqueles residentes nas regiões Norte, Nordeste, Centro-Oeste e Sul, obtiveram menor escolaridade que os adultos moradores da região Sudeste, especialmente, quando se consideram os filhos que vivem com os pais.

Tabela 2: Regressões - determinação conjunta do grau de instrução e da posição do domiciliar do filho

MQO – sem correção		MV – modelo de seleção com mudança endógena		
(1)	(2)	(3)	(4)	(5)

⁷ Dado que inexistem informações de residentes em áreas rurais da região Norte na PNAD de 1996, para testar a robustez dos resultados, as estimações foram refeitas excluindo da amostra os residentes naquela região. Conforme pode ser observado nas tabelas A.4 e A.5, no apêndice, os coeficientes de persistência estimados são bastante próximos daqueles presentes nessa seção.

	Instrução Filhos emancipados	Instrução Filhos dependentes	Instrução Filhos Emancipados	Instrução Filhos dependentes	Seleção (d=1 filho dependente)
Branco	1,2370*** (0,0396)	0,7849*** (0,0901)	1,1940*** (0,0403)	0,6543*** (0,0910)	-0,0384** (0,0170)
Idade	0,1676*** (0,0135)	-0,0218 (0,0477)	0,0477*** (0,0146)	-0,1724*** (0,0498)	-0,1493*** (0,0069)
Idade2	-0,0026*** (0,0002)	-0,0005 (0,0007)	-0,0015*** (0,0002)	0,0000 (0,0007)	0,0008*** (0,0001)
Urbana	2,5142*** (0,0383)	2,4181*** (0,0959)	2,5023*** (0,0393)	2,4183*** (0,0978)	-0,0474** (0,0212)
Educação Pai	0,6700*** (0,0053)	0,5165*** (0,0105)	0,6759*** (0,0055)	0,5472*** (0,0112)	0,0322*** (0,0021)
Norte	-0,2633*** (0,0813)	-0,5146*** (0,1819)	-0,3493*** (0,0823)	-0,5810*** (0,1820)	-0,3240*** (0,0338)
Nordeste	-0,4926*** (0,0470)	-0,9170*** (0,1011)	-0,5225*** (0,0479)	-0,9006*** (0,1022)	-0,1443*** (0,0195)
Sul	-0,0748 (0,0490)	-0,4820*** (0,1197)	-0,1699*** (0,0500)	-0,7843*** (0,1225)	-0,1939*** (0,0222)
Centro-Oeste	0,1141* (0,0626)	-0,3304** (0,1572)	-0,0368 (0,0638)	-0,7812*** (0,1630)	-0,3650*** (0,0277)
Econ. Ativo					-1,0850*** (0,0286)
Moradores					0,1793*** (0,0043)
Intercepto	-0,5173* (0,2874)	4,2028*** (0,8457)	2,9254*** (0,3198)	6,6632*** (0,8730)	3,5753*** (0,1285)
σ_1					3,8906*** (0,0365)
σ_2					3,7496*** (0,0087)
ρ_1					0,4235*** (0,0348)
ρ_2					0,4623*** (0,0201)
Observações	43.772	8.622	52.394		
R2 ajustado	0,4197	0,3447			
Teste de independência (Wald Qui-Quadrado)					632,38***

Fonte: Elaboração própria a partir dos dados da PNAD de 1996. Notas: Desvios padrão robustos à heterocedasticidade entre parênteses. *** Estatisticamente significativa a 1%. ** Estatisticamente significativa a 5%. * Estatisticamente significativa a 10%.

A despeito das evidências até então reportadas, a tabela 3, abaixo, apresenta resultados importantes para a análise da consistência entre as decisões de permanência no domicílio enquanto dependente dos pais e de investimento em capital humano. Nesse sentido, são registrados os resultados da estimação da equação estrutural (2) juntamente com os coeficientes de efeito marginal.

O resultado mais importante da tabela em análise refere-se ao sinal positivo do coeficiente δ , que mede a relação entre retorno esperado do capital humano e opção de co-residência com os pais. Conforme destacado, a opção dos filhos dependentes parece ter sido consistente com o investimento em termos de instrução acumulada. Note-se que 1 ano a mais de educação esperada eleva em 3,3 p.p a probabilidade de o filho permanecer sob a guarda dos pais quando adulto.

Tabela 3: Regressão - equação estrutural: determinação da posição do filho no domicílio na fase adulta

	(1)	(2)
	Coeficientes (d=1- filho dependente)	Efeito Marginal

Retorno esperado (educação) δ	0,1763*** (0,0287)	0,0330*** (0,0054)
Idade	-0,0666*** (0,0033)	-0,0125*** (0,0006)
Urbana	-0,0090 (0,0211)	-0,0017 (0,0040)
Educação Pai	0,0517*** (0,0045)	0,0097*** (0,0008)
Econ. Ativo	-1,0991*** (0,0285)	-0,2057*** (0,0050)
Moradores	0,1676*** (0,0040)	0,0314*** (0,0007)
Norte	-0,2855*** (0,0334)	-0,0510*** (0,0055)
Nordeste	-0,0890*** (0,0203)	-0,0171*** (0,0039)
Sul	-0,0730** (0,0311)	-0,0141** (0,0060)
Centro-Oeste	-0,2309*** (0,0337)	-0,0421*** (0,0059)
Intercepto	2,1724*** (0,0730)	
Observações	52.394	

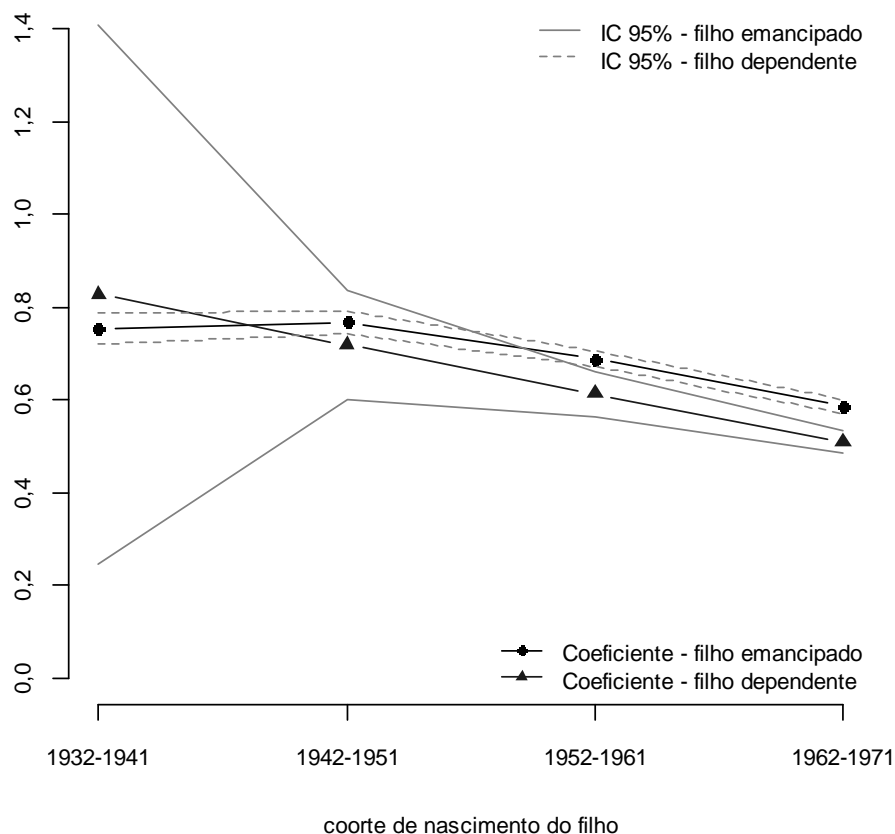
Fonte: Elaboração própria a partir dos dados da PNAD de 1996. Notas: Desvios padrão robustos à heterocedasticidade entre parênteses. *** Estatisticamente significante a 1%. ** Estatisticamente significante a 5%. * Estatisticamente significante a 10%.

Os demais resultados são consistentes com as correlações já observadas na equação estrutural reduzida. Ainda é possível observar, na tabela 3, que um ano adicional de idade diminui em 1,2 p.p a chance de o filho continuar com os pais na fase adulta, morar no meio urbano também reduz aquela chance em 0,2 p.p comparado a viver em área rural. Já para um adulto que trabalha ou procura emprego (economicamente ativo), a probabilidade de co-residir com os pais diminui em 20,6 p.p, enquanto a presença de mais um morador no domicílio aumento a chance em 3,1 p.p. Comparado a região Sudeste, residir nas regiões Norte ou Centro-Oeste produz os maiores impactos na redução da probabilidade de permanência do filho sob a guarda dos pais.

Uma questão relevante para o caso do Brasil é a investigação do comportamento da persistência intergeracional/educacional dos filhos ao longo de gerações (coortes), ou seja, procurar diferenças de persistência entre gerações antigas e recentes. Nessa perspectiva, a figura 5, a seguir, registra os resultados estimados para o coeficiente de persistência β considerando estimativas do modelo estrutural (3), (4) e (5) em diferentes amostras selecionadas por data de nascimento do filho.

Os resultados são consistentes com aqueles já demonstrados por Ferreira e Veloso (2003), ao indicarem a redução da persistência intergeracional da educação nas gerações mais recentes. Contudo, é possível observar que para as gerações nascidas a partir da década de 50, o grau de persistência dos filhos dependentes é inferior ao verificado para os filhos emancipados, por um lado, e, por outro, apresenta uma tendência de redução mais rápida quando comparada ao último grupo.

Figura 5: Coeficiente de persistência intergeracional segundo coortes de nascimento do filho e por condição no domicílio



Fonte: Elaboração própria a partir dos dados da PNAD de 1996.

Notas: Resultados obtidos pelo método *heckit* em dois estágios. Intervalos de confiança gerados por *bootstrap* com 1.000 replicações de cada regressão.

6. Conclusão

O presente estudo analisou se a ausência de co-residentes na análise de mobilidade intergeracional no Brasil produz medidas tendenciosas de persistência educacional. Em paralelo, este artigo investigou se a decisão de permanência no domicílio, ou de emancipação, está associada à expectativa de acumulação de capital humano.

A exploração inicial por medidas não paramétricas de mobilidade (matrizes de transição *markovianas*) apontou uma diferença entre a persistência dos pais menos e mais escolarizados considerando os filhos co-residentes e emancipados. Na amostra dos filhos emancipados a persistência dos pais analfabetos é maior que a dos filhos dependentes. No entanto, a persistência dos pais mais escolarizados é maior na amostra que considera dos filhos não co-residentes (emancipados). Este resultado se mantém quando se consideram as principais regiões geográficas. Contudo, entre as áreas rurais e urbanas a diferença de persistência dos pais entre as amostras por condição domiciliar dos filhos é mais acentuada. Em áreas rurais, por exemplo, a persistência dos pais analfabetos na amostra de filhos emancipados é de quase 55%, enquanto que na amostra de filhos co-residentes este valor é de 15%. Em linhas gerais, estes resultados sugerem que em áreas mais dinâmicas o efeito dos pais analfabetos tende a ser menor e menos intenso sobre os filhos co-residentes, o que se coaduna com as implicações dos modelos de Galor e Zeira (1993) e Galor e Tsiddon (1997).

Os resultados do modelo estrutural reforçam a conclusão da análise não paramétrica. Primeiramente, os coeficientes que indicam a persistência educacional são significativamente diferentes considerando a amostra de filhos emancipados e não emancipados. A persistência

educacional é maior para os filhos não dependentes se comparado aos co-residentes. Deste modo, é um erro generalizar os coeficientes considerando apenas os filhos emancipados.

Ainda considerando o modelo paramétrico estrutural é possível afirmar a decisão de permanecer com os pais está correlacionada com a acumulação de capital humano dos filhos. Os achados sugerem que 1 ano a mais de educação esperada eleva em 3,3 p.p a probabilidade de o filho permanecer sob a guarda dos pais quando adulto. Neste contexto, ainda vale observar que a maior idade do filho, a condição de atividade econômica, morar nas regiões Norte e Centro-Oeste e/ou em áreas urbanas são características que impactam negativamente (positivamente) quando na chance de o filho co-residir com o pai (emancipado).

Por fim, tendo como base a análise por coortes de idade de nascimento dos filhos, percebeu-se que as gerações nascidas a partir da década de 50 apresentam um grau de persistência dos filhos dependentes inferior ao verificado para os filhos emancipados. Contudo, verificou-se que esta geração, os filhos dependentes apresentam uma tendência de redução mais rápida quando comparada aos emancipados.

7. Referências

- ATKINSON, A. B. On Intergenerational Income Mobility. *Journal of Post Keynesian Economics*. v. 3, 1981.
- BARROS, R., LAM, D. Desigualdade de renda, desigualdade em educação e escolaridade das crianças no Brasil. *Pesquisa e Planejamento Econômico*, v.23, n.2, 1993.
- BARROS, R., MENDONÇA, R., SANTOS, D., QUINTAES, G. Determinantes do desempenho educacional no Brasil. *Pesquisa e Planejamento Econômico*, v.31, n.1, Abr. 2001.
- BECKER, G. S., MURPHY, K. M., TAMURA, R. Human Capital, Fertility, and Economic Growth. *Journal of Political Economy*, 98, S12-S37, 1990.
- BEHRMAN, J., GARVIRIA, A., SZÉKELI, M.. Intergenerational mobility in Latin America. *Economia*, v. 2, n. 1, 2001.
- BORJAS, G. Ethnic capital and intergenerational mobility. *The Quarterly Journal of Economics*, v. 107, n. 1, 1992.
- CAMERON, A. C., TRIVEDI, P. K. *Microeconometrics: methods and applications*. Cambridge University Press, New York, 2005
- COUCH, K., DUNN, T. Intergenerational correlations in labor market status: a comparison of the United States and Germany. *Journal of Human Resources*, v. 32, n. 1, 1997.
- ERMISCH, J., FRANCESCONI, M. Family Matters: Impacts of Family Background on Educational Attainments. *Economica*. v. 68, 2001.
- ERMISCH, J., NICOLETTI. Intergenerational Earnings Mobility: changing across cohorts in Britain. *The B. E. Journal of Economic Analysis & Policy*, v. 7, n. 2, 2006.
- FERREIRA, S.; VELOSO, F.. Mobilidade intergeracional de educação no Brasil. *Pesquisa e Planejamento Econômico*, Rio de Janeiro, v. 33, n. 3, 2003.
- FIGUEIRÊDO, E. A. de; NETTO JR., J. L. da; PÔRTO JR., S. da S. Distribuição, mobilidade e polarização de renda no Brasil: 1987 a 2003. *Revista Brasileira de Economia*, v. 61, n. 1, 2007.

- FIELDS, G. S., OK, E. A. The Meaning and Measurement of Income Mobility. *Journal of Economic Theory*. Elsevier, vol. 71(2), 1996.
- GALOR, O., TSIDDON, D. Technological progress, mobility and economic growth. *The American Economic Review*, v. 87, 1997.
- GALOR, O., ZEIRA, J. Income distribution and macroeconomics. *Review of Economics Studies*, v. 60, 1993.
- HALL, B. H. Notes on sample selection models. Disponível em <<http://elsa.berkeley.edu/~bhhall/e244/sampsel.pdf>>, 2002.
- HECKMAN, J. Sample selection bias as a specification error. *Econometrica*, v.47, n.1, 1979.
- LEE, L. Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables. *International Economic Review*, v.19, n. 2, 1978.
- LOKSHIN, M., SAJAIA, Z. Maximum likelihood estimation of endogenous switching regression models. *The Stata Journal*, v.4, n. 3, 2004.
- LUCAS, R. E. On The Mechanics of Economic Development. *Journal of Monetary Economics*, v. 22, 1988.
- MACHADO, D. C., GONZAGA, G. Impactos dos Fatores Familiares Sobre a Defasagem Idade-Série de Crianças no Brasil. *Revista Brasileira de Economia*, v. 61, n.4, 2007.
- MADDALA, G. S. *Limited-Dependent and Qualitative Variables in Economics*. New York: Cambridge University Press, 1983.
- MULLIGAN, C. *Parental priorities and economic inequality*. Chicago: University of Chicago Press, 1997.
- NETTO JR., J. L. da S.; PÔRTO JR., S. da S.; FIGUEIREDO, E. A. de. Migração e Distribuição de Capital Humano no Brasil: mobilidade intergeracional educacional e intrageracional de renda. *Revista Econômica do Nordeste*, v. 39, p. 404-427, 2008.
- PÈRO, V., SZERMAN, D. Mobilidade Intergeracional de Renda no Brasil. *Pesquisa e Planejamento Econômico*. v. 38, n.1, abr. 2008.

Apêndices

(A)

Seguindo Maddala (1983), é possível mostrar que o valor esperado da escolaridade do filho, condicionado a escolha de cada regime domiciliar é dado por:

$$E(Y_{1i}^c | d = 1) = \eta_1 + E(\varepsilon_{1i} | \varepsilon_i \geq -V_i) = \eta_1 + \hat{\sigma}_1 \hat{\rho}_{1\varepsilon} \frac{f(\hat{V}_i)}{F(\hat{V}_i)} \quad (8)$$

$$E(Y_{2i}^c | d = 0) = \eta_2 + E(\varepsilon_{2i} | \varepsilon_i < -V_i) = \eta_2 - \hat{\sigma}_2 \hat{\rho}_{2\varepsilon} \frac{f(\hat{V}_i)}{1 - F(\hat{V}_i)} \quad (9)$$

Onde: η_1 e η_2 representam valores médios dos atributos observados e parâmetros corrigidos para viés de seleção, \hat{V}_i é a predição linear da equação (5), f é a função de densidade normal e F a função de densidade acumulada. Em particular, os termos $\frac{f(\hat{\mu}Z_i)}{F(\hat{\mu}Z_i)}$ e $\frac{f(\hat{\mu}Z_i)}{1-F(\hat{\mu}Z_i)}$ são medidas de probabilidade de seleção amostral, conhecidas como razões inversas de Mills, ambas positivas.

Destarte, o sentido do viés de seleção provocado pela condição do filho do domicílio dependerá dos sinais dos coeficientes de correlação $\hat{\rho}_{1\varepsilon}$ e $\hat{\rho}_{2\varepsilon}$. No caso de $\hat{\rho}_{1\varepsilon} > 0$ ($\hat{\rho}_{2\varepsilon} < 0$), os filhos situados no regime 1 (regime 2) têm escolaridade média superior aquela verificada para um indivíduo aleatoriamente escolhido na amostra, ou seja, são positivamente selecionados em atributos não observados. Já a magnitude do viés de seleção depende: (i) do tamanho de $\hat{\rho}_{1\varepsilon}$ e $\hat{\rho}_{2\varepsilon}$ (ii) do tamanho do desvio padrão σ_1 (σ_2) e da severidade do ponto de truncagem da distribuição normal, isto é, $\frac{f(\hat{V}_i)}{F(\hat{V}_i)}$ será maior quanto menor for o valor limiar \hat{V}_i (MADDALA, 1983 ; HALL, 2002).

(B)**Tabela A.1: Descrição das variáveis usadas nas regressões**

Variável	Definição
Educação Filho	(V4703) - Anos de estudo
Branco	(V0404) 1 - se o filho declara-se de cor branca e 0 - caso contrário*.
Idade	(V8005) Idade do filho em anos.
Idade2	(V8005) Idade do filho em anos elevada ao quadrado (experiência).
Urbana	(V4728) 1 - se o filho reside em zona urbana e 0 - em zona rural*.
Educação Pai	(V4703; V1217) Nível de instrução do pai. No caso dos filhos emancipados, recebeu o valor 0, caso o pai tenha menos de 1 ano de estudo; 2 caso tenha completado a primeira, segunda ou terceira série do ensino fundamental ou quarta série incompleta; 4 se completou a quarta série; 6 se concluiu a quinta, sexta ou sétima série, mesmo sem completar a oitava série; 8 se concluiu a oitava série; 10 caso possua ensino médio não concluído; 11 se completou o ensino médio; 13 se cursou algum curso superior e não concluiu o mesmo ou tem 15 anos de escolaridade sem completar curso superior; 16 caso tenha curso superior completo ou frequentado pós-graduação. Já no caso dos filhos co-residentes, a compatibilização procedeu-se da seguinte forma: o nível de instrução do pai recebeu o valor 0, caso o pai tenha registrado menos de 1 ano de estudo; 2 caso tenha de 1 a 3 anos completos de estudo; 4 se completou 4 anos de estudo; 6 se possui de 5 a 7 anos de estudo; 8 se registra 8 anos de estudo; 10 se tem de 9 a 10 anos de estudo; 11 caso tenha 11 anos de estudo; 13 se registra de 12 a 15 anos de estudo e sem completar curso superior; 16 se tem 15 ou mais anos de estudo tendo completado o curso superior ou frequentado pós-graduação.
Econ. Ativo	(V4704) 1 - se o filho era economicamente ativo na semana de referência e 0 - caso contrário*.
Moradores	(V0115) - Número de moradores no domicílio.
Norte	(UF) 1 - se o filho reside na região Norte e 0 - se reside na região Sudeste*
Nordeste	(UF) 1 - se o filho reside na região Nordeste e 0 - se reside na região Sudeste*
Sul	(UF) 1 - se o filho reside na região Sul e 0 - se reside na região Sudeste*
Centro-Oeste	(UF) 1 - se o filho reside na região Centro - Oeste e 0 - se reside na região Sudeste*
Tipo Filho	(V0401) 1 - se a pessoa ocupa a posição de filho no domicílio - 0 - se a pessoa ocupa a posição de chefe ou cônjuge no domicílio*

Fonte: Elaboração própria a partir dos dados da PNAD de 1996. Notas: * Categoria omitida. Entre parênteses, encontra-se a variável de origem do dicionário da PNAD.

Tabela A.2: Matriz de transição – filhos emancipados - %

Escolaridade do pai	Escolaridade do filho								
	0	2	4	6	8	10	11	13	16
0	33,85	23,72	18,49	10,71	5,65	1,65	4,20	0,65	1,08
2	9,03	19,15	22,43	17,48	11,36	3,19	11,37	2,04	3,96
4	2,78	5,86	15,67	15,45	15,15	5,95	22,00	5,54	11,59
6	1,40	5,47	6,58	17,25	13,17	8,50	25,81	7,67	14,15
8	1,38	2,44	4,07	8,68	13,71	6,07	28,78	10,43	24,44
10	0,00	1,29	1,66	8,56	8,53	7,49	31,95	9,65	30,86
11	0,37	1,20	1,76	5,09	6,48	5,06	32,56	11,68	35,80
13	0,00	1,47	3,03	4,72	9,69	3,12	25,88	13,26	38,82
16	0,75	0,67	0,90	2,74	3,77	1,99	16,19	12,98	60,02

Tabela A.3: Matriz de transição – filhos dependentes - %

Escolaridade do pai	Escolaridade do filho								
	0	2	4	6	8	10	11	13	16
0	33,85	23,72	18,49	10,71	5,65	1,65	4,20	0,65	1,08
2	9,03	19,15	22,43	17,48	11,36	3,19	11,37	2,04	3,96
4	2,78	5,86	15,67	15,45	15,15	5,95	22,00	5,54	11,59
6	1,40	5,47	6,58	17,25	13,17	8,50	25,81	7,67	14,15
8	1,38	2,44	4,07	8,68	13,71	6,07	28,78	10,43	24,44
10	0,00	1,29	1,66	8,56	8,53	7,49	31,95	9,65	30,86
11	0,37	1,20	1,76	5,09	6,48	5,06	32,56	11,68	35,80
13	0,00	1,47	3,03	4,72	9,69	3,12	25,88	13,26	38,82
16	0,75	0,67	0,90	2,74	3,77	1,99	16,19	12,98	60,02

Tabela A.4: Regressões - determinação conjunta do grau de instrução e da posição do domiciliar do filho – sem moradores da região Norte

	MQO – sem correção		MV – modelo de seleção com mudança endógena		
	(1) Instrução Filhos emancipados	(2) Instrução Filhos dependentes	(3) Instrução Filhos emancipados	(4) Instrução Filhos dependentes	(5) Seleção (d=1 filho dependente)
Branco	1,2289*** (0,0406)	0,8190*** (0,0928)	1,1889*** (0,0414)	0,6800*** (0,0939)	-0,0378** (0,0176)
Idade	0,1687*** (0,0139)	-0,0236 (0,0492)	0,0500*** (0,0150)	-0,1731*** (0,0514)	-0,1494*** (0,0071)
Idade2	-0,0026*** (0,0002)	-0,0004 (0,0007)	-0,0015*** (0,0002)	0,0001 (0,0007)	0,0008*** (0,0001)
Urbana	2,5243*** (0,0388)	2,4055*** (0,0970)	2,5085*** (0,0398)	2,3989*** (0,0990)	-0,0539** (0,0215)
Educação Pai	0,6708*** (0,0055)	0,5194*** (0,0108)	0,6770*** (0,0057)	0,5513*** (0,0116)	0,0323*** (0,0022)
Nordeste	-0,4923*** (0,0471)	-0,9041*** (0,1013)	-0,5213*** (0,0479)	-0,8892*** (0,1024)	-0,1427*** (0,0196)
Sul	-0,0713 (0,0491)	-0,4918*** (0,1199)	-0,1657*** (0,0500)	-0,7971*** (0,1231)	-0,1958*** (0,0223)
Centro-Oeste	0,1166* (0,0627)	-0,3218** (0,1573)	-0,0317 (0,0638)	-0,7803*** (0,1638)	-0,3651*** (0,0277)
Econ. Ativo					-1,0862*** (0,0297)
Moradores					0,1760*** (0,0045)
Intercepto	-0,5779* (0,2951)	4,1758*** (0,8717)	2,8267*** (0,3283)	6,6164*** (0,9009)	3,5967*** (0,1316)
σ_1					3,8840*** (0,0403)
σ_2					3,7356*** (0,0087)
ρ_1					0,4308*** (0,0374)
ρ_2					0,4550*** (0,0206)
Observações	41.161	8.078		49.239	
R2 ajustado	0,4232	0,3536			
Teste de independência (Wald Qui-Quadrado)					573,19***

Fonte: Elaboração própria a partir dos dados da PNAD de 1996. Notas: Desvios padrão robustos à heterocedasticidade entre parênteses. *** Estatisticamente significante a 1%. ** Estatisticamente significante a 5%. * Estatisticamente significante a 10%.

Tabela A.5: Regressão - equação estrutural: determinação da posição do filho no domicílio na fase adulta - sem moradores da região Norte

	(1)	(2)
	Coefficientes (d=1- filho dependente)	Efeito Marginal
Retorno esperado (educação) δ	0,1909*** (0,0308)	0,0357*** (0,0058)
Idade	-0,0649*** (0,0036)	-0,0122*** (0,0006)
Urbana	-0,0103 (0,0215)	-0,0019 (0,0040)
Educação Pai	0,0535*** (0,0047)	0,0100*** (0,0009)
Econ. Ativo	-1,0948*** (0,0296)	-0,2050*** (0,0051)
Moradores	0,1651*** (0,0042)	0,0309*** (0,0007)
Nordeste	-0,0850*** (0,0204)	-0,0161*** (0,0039)
Sul	-0,0616* (0,0325)	-0,0118* (0,0062)
Centro-Oeste	-0,2198*** (0,0347)	-0,0397*** (0,0061)
Intercepto	2,1599*** (0,0774)	
Observações	49.239	

Fonte: Elaboração própria a partir dos dados da PNAD de 1996. Notas: Desvios padrão robustos à heterocedasticidade entre parênteses. *** Estatisticamente significante a 1%. ** Estatisticamente significante a 5%. * Estatisticamente significante a 10%.