

# Uma Análise da Transmissão Intergeracional de Capital Humano no Brasil

Marcio Gold Firmo  
World Bank

Rodrigo R. Soares  
PUC-Rio

## Resumo

Neste trabalho, avaliamos a transmissão intergeracional de capital humano no Brasil, separando o efeito da escolaridade de mãe e pai sobre a defasagem idade-série de filhos e filhas na escola. A partir de dados das PNADs de 1988 e 1996, estimativas através do método de Mínimos Quadrados Ordinários indicam presença de fortes não-linearidades nessas relações e efeitos diferentes para meninos e meninas. Utilizando séries históricas de escolas e professores por Unidade da Federação do IBGE como fonte de identificação do efeito causal de escolaridade dos pais sobre o desempenho dos filhos, o método de Mínimos Quadrados em Dois Estágios com variáveis instrumentais aponta efeitos significantes tanto de escolaridade da mãe quanto do pai, mas não permite distingui-los adequadamente.

Palavras-chave: educação, transmissão intergeracional, efeito causal, efeito-gênero

## Abstract

In this paper we assess the intergenerational transmission of human capital in Brazil, separating mother's and father's effects on son's and daughter's school performance. Using 1988 and 1996 PNAD data, our OLS estimates of these relations indicate strong non-linearities and gender effects. Taking advantage of IBGE's historical series of schools and teachers in Brazilian states over the XXth century, we try to identify the causal effect of parent's education on children's school performance. The 2SLS-IV estimates point to significant effects of both mother's and father's education, but the strategy does not allow us to properly disentangle those effects.

Key words: education, intergenerational transmission, causal effect, gender effect

Classificação JEL: I20, I21, O15

Area ANPEC: Area 11 - Economia Social e Demografia Econômica

# Uma Análise da Transmissão Intergeracional de Capital Humano no Brasil

## 1 Introdução

O objetivo deste trabalho é estudar o processo de transmissão de capital humano entre gerações no Brasil, buscando inferir a existência de um efeito causal direto de escolaridade dos pais em desempenho escolar dos filhos. A existência e a magnitude desse efeito têm conseqüências sobre a persistência intergeracional de capital humano, e sua correta mensuração pode ajudar a desenvolver políticas educacionais mais eficientes.

Essa persistência tem sido objeto freqüente de investigação na literatura internacional, assim como suas possíveis causas: o arranjo institucional e cultural em que a sociedade está disposta; a transmissão entre membros de uma mesma família de atributos que poderiam estar implicitamente determinando o quanto uma pessoa irá estudar; e a própria influência que educação dos pais pode ter na educação dos filhos. O efeito de cada uma dessas causas é ainda em grande parte desconhecido.

No caso da influência da educação dos pais sobre a dos filhos, é possível distinguir dois tipos de efeitos: os *indiretos*, através dos quais a educação de certa geração de uma família se reflete em variáveis que irão por sua vez influenciar a educação dos filhos (como no caso de um pai que, por ter se educado mais, irá ganhar mais e poderá pagar uma universidade para seu filho); e os *diretos*, que implicariam uma relação direta de causalidade em que a educação dos pais geraria através de fatores potencialmente não-observáveis, como ambiente familiar e escolha do tipo de educação que será dada aos filhos, uma maior escolaridade dos filhos. A esse efeito direto chamamos *efeito causal* da educação dos pais na educação dos filhos.

Embora os efeitos indiretos há muito sejam alvo de estudo, os efeitos diretos da educação não tem a mesma participação na literatura. Por sua dificuldade de identificação e mensuração, a maioria dos estudos data de pouco tempo atrás e suas conclusões ainda são incipientes. A correta compreensão da medida desses efeitos é fundamental para a análise mais adequada das correlações intergeracionais de capital humano e riqueza observadas nas sociedades. Além disso, a literatura vem recentemente destacando a importância da análise separada de efeitos de mãe e pai sobre filhos e filhas, considerando a possibilidade de haver efeitos distintos para cada relação entre os gêneros (*gender effects*). Destacam-se Becker e Tomes (1986), Charles e Hurst (2003), Ferreira e Veloso (2003), Checci e Flabbi (2007), Black *et al.* (2005), Oreopoulos *et al.* (2006), Emerson e Souza (2006), Souza (2007) e Machado e Gonzaga (2007), trabalho que mais se aproxima do nosso.

A evidência de efeito causal da educação dos pais na educação dos filhos ainda é bastante esparsa e inconclusiva. A maioria dos trabalhos que se dedicaram a isolá-los dos outros efeitos familiares encontra alguma evidência de efeito materno, e muito pouca evidência de efeito paterno. As magnitudes dos efeitos encontrados são bastante pequenas, mas há razões para crer que os efeitos podem ser muito maiores em ambientes de países em desenvolvimento que em ambientes de países já desenvolvidos. A existência de não-linearidades e possíveis relações mais fortes para níveis mais baixos de escolaridade poderiam explicar os pequenos e pouco significantes efeitos em contextos de sociedades desenvolvidas sem que o mesmo esteja ocorrendo em sociedades mais pobres. A produção de estimativas comparáveis com as existentes na literatura internacional pode, então, esclarecer um pouco mais os processos de transmissão intergeracional de capital humano.

O presente trabalho analisa os padrões de correlação intergeracional de educação no Brasil e busca evidências de um efeito causal da educação dos pais na educação dos filhos, separando os efeitos de mãe e pai sobre filhos e filhas. Analisaremos os dados das PNADs de 1996 e 1988, juntamente com um conjunto de variáveis que, sob as hipóteses de identificação propostas, poderão nos ajudar a isolar o efeito causal desejado de outros efeitos.

Os principais resultados sugerem que educação da mãe e do pai afetam de forma significativa o desempenho escolar de seus filhos através de efeitos que vão além de canais de renda e riqueza. Encontram-se indícios de componentes de gênero nessas relações: mães parecem influenciar mais os filhos que pais, e meninos são mais influenciados que meninas. A persistência educacional é maior nas camadas mais pobres da população, e a evidência de não-linearidades é bastante clara, implicando uma relação côncava entre educação dos pais e dos filhos. Analisados separadamente, tanto educação da mãe

quanto do pai parecem ter efeito causal significativo sobre o desempenho dos filhos, mas a identificação do efeito causal parcial de cada um é prejudicada pela alta correlação entre essas variáveis, assim como entre seus instrumentos. Analisados conjuntamente, encontramos forte evidência de efeito causal do capital humano dos pais tanto em seus filhos quanto em suas filhas.

O capítulo seguinte descreve o modelo teórico e a estratégia empírica utilizada, e o capítulo 3 descreve os dados que formam nossa base. No capítulo 4 desenvolvemos uma análise por Mínimos Quadrados Ordinários (MQO) em que, com a sucessiva inclusão de variáveis de controle, boa parte da endogeneidade e do viés de variáveis omitidas não-observadas pode ser expurgado. O capítulo 5 desenvolve a análise por 2SLS, mostrando como os instrumentos se relacionam com as variáveis endógenas e quais os resultados das instrumentalizações para as variáveis de interesse, discutindo também os possíveis problemas encontrados. O capítulo 6 conclui.

## 2 Estratégia Empírica

Na busca pelos efeitos da educação dos pais na educação dos filhos, utilizamos duas estratégias. Primeiro, estimamos esse efeito através do método de Mínimos Quadrados Ordinários, inserindo progressivamente controles que isolam os efeitos procurados de outros fatores não-observados que estejam influenciando a relação estimada. Em uma seção seguinte, procuramos estimar o real efeito causal da educação dos pais nos filhos, através do método de Mínimos Quadrados em Dois Estágios (2SLS) com variáveis instrumentais. Para tal, devemos ter claro o modelo gerador de educação dos filhos, as limitações a que estamos sujeitos nessa análise empírica, e as potenciais formas de superá-las.

### 2.1. O Modelo

Utilizamos um modelo de transmissão intergeracional de educação *à la* Becker e Tomes (1986), com algumas peculiaridades. O capital humano de uma família se relaciona da seguinte forma:

$$ed_{filho}^i = \beta * ed_{pais}^i + \delta * X^i + u_{filho}^i \quad (1)$$

Onde  $ed_{filho}^i$  e  $ed_{pais}^i$  são, respectivamente, a educação de um indivíduo da geração da família  $i$  e seus pais,  $X$  é um vetor de características observáveis de filho, mãe e pai, e  $u_{filho}^i$  é um erro aleatório não-observável.

Portanto, o capital humano do filho depende de fatores observáveis, como sua idade e a renda de seus pais, mas também diretamente do capital humano de pai e de mãe. É o caso de complementaridade entre educação dos pais e dos filhos na função de produção de educação da criança. Nesse sentido é que chamamos  $\beta$  o efeito causal da educação de uma geração sobre a outra.

O erro  $u$  contém fatores que influenciam o nível de capital humano do filho. Como enfatiza a literatura, há um conjunto enorme de fatores que podem estar exercendo esse papel. Basicamente, podemos dividi-los em três grupos: (i) fatores familiares, passados de geração a geração, como cultura familiar, preferência por trabalho e/ou estudo, talento e habilidade ou inteligência; (ii) fatores comuns a todos em um dado período de tempo ou certo lugar, (iii) atributos exclusivamente pessoais do indivíduo.

Matematicamente:

$$u_{filho}^i = v_{filho}^i + w^i + \varepsilon_{filho}^i \quad (2)$$

Em que  $v_{filho}^i$ ,  $w^i$  e  $\varepsilon_{filho}^i$  correspondem respectivamente a (i), (ii) e (iii) do parágrafo anterior. As hipóteses que se fazem são de que o termo  $\varepsilon_{filho}^i$  é um erro estocástico não-correlacionado com qualquer outra variável e de que os fatores familiares são transmitidos da seguinte forma:

$$v_{filho}^i = \rho * v_{pais}^i, \quad 0 < \rho < 1 \quad (3)$$

A partir das três equações acima fica claro que  $Cov(ed_{pais}^i, u_{filho}^i) \neq 0$ . Assim, ao fazermos a projeção de  $ed_{filho}^i$  em  $ed_{pais}^i$ , estaremos captando não apenas o efeito causal deste naquele, mas uma correlação entre essas variáveis que incluirá (i) o efeito dos fatores familiares não-observáveis que influenciam educação de pais e filhos simultaneamente; (ii) os efeitos dos canais através dos quais quaisquer outras variáveis não incluídas entre os controles (como a renda dos pais, por exemplo) influenciem o capital humano do filho. É isso que motiva o estudo mais cuidadoso dessas relações, através tanto de modelos MQO mais bem especificados quanto da busca por variáveis instrumentais que

sejam capazes de isolar esse efeito causal. Como vamos distinguir entre os efeitos de mãe e pai, o modelo estimado é um pouco mais abrangente. As equações são as seguintes:

$$ed_{filho}^i = \alpha_j * ed_{mãe}^i + \beta_j * ed_{pai}^i + c * X^i + u_{filho}^i, \quad j = h, m \quad (4)$$

$$u_{filho}^i = \lambda * u_{mãe}^i + \phi * u_{pai}^i + \varepsilon_{filho}^i, \quad 0 < \lambda < 1, \quad 0 < \phi < 1 \quad (5)$$

Em caso de *matching* no mercado de casamentos, temos ainda:

$$Cov(u_{mãe}, u_{pai}) \neq 0 \quad (6)$$

Em que  $\alpha_j$  e  $\beta_j$  são respectivamente o efeito da educação da mãe e do pai no filho de sexo  $j$ , o distúrbio  $u$  é influenciado tanto pelo distúrbio materno quanto paterno, e  $\varepsilon_{filho}^i$  é o distúrbio não-correlacionado.

## 2.2. As Estimações por MQO

O primeiro conjunto de regressões busca identificar padrões na correlação entre os níveis de capital humano observados. Idealmente, gostaríamos de uma medida do capital humano final atingido pelo filho. No entanto, a PNAD não possui esses dados, uma vez que os filhos que têm seus dados coletados são aqueles que ainda vivem com os pais e, em sua grande maioria, têm menos que 18 anos. A solução encontrada é seguir uma estratégia parecida com Horowitz e Souza (2004), criando uma medida de defasagem idade-série *normalizada* que, sob hipóteses bastante gerais, é uma *proxy* para o nível de capital humano atingido no futuro.

O Ministério da Educação (MEC) determina que toda criança com 7 anos completos até 1º de março deve estar matriculada na 1ª série. Se a criança satisfizer esse requisito e não for reprovada em nenhum ano, quando ela tiver 7+y anos completos até 1º de março ela estará na y+1-ésima série. A medida de defasagem idade-série *bruta* é simplesmente quantos anos de estudo completos a menos tem uma criança em relação àquela com desempenho ideal. Por ser uma variável limitada cujo conjunto de valores possíveis se desloca com o avanço da idade da criança, devemos fazer uma normalização para podermos utilizar o modelo clássico de regressão linear. Nossa variável dependente será:

$$y_i = \left( \frac{defas_i}{idade_i - 7} \right) \times 100$$

Em que  $defas_i$  é defasagem idade-série bruta do filho 'i' e  $idade_i$  é a idade escolar, ou seja, sua idade em 1º de março do ano em que a pesquisa foi feita. Um valor de  $y$  igual a zero representa um filho com percurso escolar ideal, sem atrasos. Um valor igual a 100 representa um filho *totalmente atrasado*. Desse modo, podemos pensar nossa variável como uma medida do percentual de atraso da criança em relação ao desempenho ideal. Com essa variável estamos capturando o efeito da educação dos pais em (i) a probabilidade de o filho ser reprovado 'vezes' o número de anos que ele já está na escola, (ii) a chance de o filho largar a escola antes de completar 18 anos e (iii) o filho entrar atrasado na escola *desde o início*.

Apesar de estarmos englobando três efeitos diferentes em uma mesma variável, acreditamos que todos estarão correlacionados com o nível final de capital humano atingido pelos filhos, que é o que gostaríamos de avaliar. Além disso, dada a enorme porcentagem de alunos no Brasil que têm defasagem idade-série estritamente positiva, essa variável tem vantagens sobre aquela utilizada por Machado e Gonzaga (2007), nos permitindo distinguir e quantificar os casos mais adequadamente, com também avaliar possíveis não-linearidades existentes. Apesar da falta de modelos satisfatórios que impliquem diferentes efeitos tanto de pai *versus* mãe quanto sobre meninos *versus* meninas, uma crescente parte da literatura empírica tem abordado essa questão, com bons indícios da existência dessas diferenças. Portanto, a princípio estimaremos separadamente as relações mãe-filho, mãe-filha, pai-filho e pai-filha.

Nossa especificação básica inclui como controles a idade escolar, idade escolar ao quadrado, idade e idade ao quadrado de mãe e pai, além de *dummies* que controlam para efeitos fixos de: gênero e cor do filho e Unidade da Federação de nascimento de filhos, mães e pais.

Em seguida, incluímos entre os controles a renda per capita familiar para ver o quanto da transmissão intergeracional de educação estimada inicialmente está se dando *através* da renda.

É mais provável que o investimento em capital humano esteja sendo influenciado por uma variável de estoque de riqueza, e não de fluxo. Para tratar do caso em que a medida relevante seja a renda

permanente, construímos variáveis indicadoras de características domiciliares e as incluímos numa 3ª especificação. Por último, adicionamos como controles variáveis que indicam o nível educacional atingido por avós paternos e maternos, que podem conter informação sobre (i) renda permanente familiar; (ii) características familiares não-observáveis. A essa especificação damos o nome “completa”. Em todas as estimações permitimos *clustering* do erro padrão no nível da família.

Por mais que as equações estimadas sejam bem especificadas, a existência de fatores familiares não-observáveis faz com que as estimativas MQO sejam viesadas. De (1), temos:

$$ed_{filhos}^i = \beta * ed_{pais}^i + \delta * X^i + u_{filhos}^i$$

$$ed_{pais}^i = \beta * ed_{avós}^i + \delta * X^i + u_{pais}^i$$

Por simplicidade, vamos ignorar os  $X^i$ . Lembrando de (2) e (3), então:

$$\hat{\beta}_{MQO} = \beta + \rho * Cov(ed_{pais}, u_{pais}) / Var(ed_{pais})$$

Logo, ainda que incluíssemos todas as características observáveis relevantes e perfeitamente medidas, o estimador de  $\beta$  por MQO ainda possuiria um viés positivo.

### 2.3. A Busca pelo Efeito Causal: a Estratégia de Variáveis Instrumentais

Como explicado acima, não se pode afirmar que os estimadores MQO encontrados sejam evidência de uma causalidade direta como aquela encontrada em Black *et al.* (2005) e Oreopoulos *et al.* (2006). Para isolar os efeitos de simultaneidade e hereditariedade existentes devido à correlação intergeracional de habilidade, talento, cultura ou outros, recorreremos à utilização de variáveis instrumentais.

Um bom instrumento tem que satisfazer dois requisitos: (i) ser não-correlacionado com o erro da equação estimada; (ii) ser suficientemente correlacionado com a variável que se supõe endógena. O segundo requisito é observável através de testes estatísticos simples; o primeiro tem como principal suporte a argumentação teórica.

O grupo de variáveis que utilizaremos como instrumentos já foi usado em Emerson e Souza (2007) e Machado e Gonzaga (2007). Trata-se do número de escolas por habitante e do número de professores por escola na Unidade da Federação em que o pai (ou a mãe) nasceu, no ano de seu nascimento.

Sabemos que a oferta de serviços educacionais no Brasil durante o século XX foi bastante precária na maior parte do país. Supomos que variações nessas séries em certa UF em dado ano representam variações na oferta de educação para as crianças que estavam em idade escolar naquela UF àquela época. Essas variações na oferta educacional devem estar relacionadas a variações no custo e no benefício de se educar. Devem, portanto, estar correlacionadas com o nível de educação adquiridos por essas crianças. Além disso, parece razoável supor que essas variações *não* estão correlacionadas com variações das características não-observáveis de cada família ou indivíduo que poderiam estar gerando os problemas de simultaneidade e viesando os coeficientes estimados. Se isso for verdade, então essas variáveis podem constituir bons instrumentos para a educação dos pais, e o método de 2SLS pode nos dar os reais coeficientes  $\alpha_j$  e  $\beta_j$  da equação (4).

Embora a hipótese de que essas variações nas ofertas de escolas e de professores sejam exógenas aos níveis familiar e individual pareça coerente, devemos fazer uma análise cuidadosa para que nossos instrumentos estejam capturando o efeito desejado, que é o de variações exógenas na oferta de serviços escolares. Por se tratarem de séries históricas anuais, também poderiam estar indiretamente caracterizando melhorias ocorrendo em várias outras dimensões da provisão de bens públicos, possivelmente muito correlacionadas com as coortes de nascimentos dos indivíduos. Essa questão será tratada no capítulo 5.

## 3 Dados

Os dados utilizados nessa análise se subdividem em dois grupos. O primeiro grupo é composto pelas Pesquisa Nacional por Amostragem a Domicílios realizadas em 1988 e 1996 (doravante PNAD88 e PNAD96, respectivamente). O segundo são o número de escolas, professores e população por Unidade da Federação durante o século XX.

### 3.1.A Pesquisa Nacional por Amostragem a Domicílios

A PNAD é uma pesquisa feita anualmente desde o fim dos anos 60 no Brasil. A partir dos anos 80 passou a conter uma amostra nacionalmente representativa de características dos domicílios e seus

residentes, com exceção da área rural da Região Norte. As PNADs de 1988 e 1996 cobrem cerca de cem mil domicílios e recolhem dados demográficos e de educação de mais de trezentos mil habitantes. Especificamente nesses dois anos, foram colhidos dados sobre o nível educacional dos pais do chefe da família e do cônjuge que nos permitem analisar com um pouco mais de precisão o processo de transmissão intergeracional de educação.

Como utilizaremos a defasagem idade-série dos filhos, nossa unidade de observação são crianças de 7 a 18 anos. Para corretamente inferir a influência parcial de pai e mãe na educação dos filhos, optamos por restringir a amostra a famílias completas, isto é, aquelas compostas por pai e mãe.

Além de variáveis de renda da família, a PNAD também coleta informações sobre certas características do domicílio em que seus membros habitam, tais como acesso a água encanada e eletricidade, presença de banheiro na parte interna do domicílio e disponibilidade de eletrodomésticos (geladeira, fogão, telefone, e outros). Com isso, podemos construir certas variáveis categóricas que podem ser utilizadas como *proxies* para a riqueza ou a “renda permanente”, variável que se esperaria estar realmente influenciando as decisões educacionais. A Tabela 1 enumera as variáveis utilizadas:

defasagem idade-série	número de anos de atraso escolar de um(a) filho(a) dividido pelo número de anos máximo que ele(a) poderia estar atrasado(a)
anos de estudo da mãe (e quadrado)	número de anos de estudo completos da mãe
anos de estudo do pai (e quadrado)	número de anos de estudo completos do pai
idade em fevereiro e termo quadrático	captura a tendência à reprovação dos filhos à medida que se tornam mais velhos
ser homem	dummy que captura possível diferença em níveis de defasagem entre meninos e meninas
ser branco	dummy que captura a diferença entre brancos e não-brancos
ln da renda familiar per capita	efeito da renda da família sobre o desempenho escolar dos filhos
idade do pai e da mãe e termos quadráticos	isola o efeito que idade dos pais pode ter sobre o desempenho dos filhos
número de irmãos/irmãs	captura efeitos do número de irmãos/irmãs sobre o desempenho
dummies de características do domicílio (PNAD96: água encanada, banheiro, telefone, geladeira, freezer, máquina de lavar) (PNAD88: domicílio permanente, eletricidade, geladeira)	controla parcialmente por riqueza ausente da variável de renda familiar
dummies de escolaridade dos avós (dez dummies para cada avô, representando desde "sem escolaridade" até "pós-graduação completa")	controla parcialmente por fatores familiares não-observados e <i>background</i> familiar
dummies da UF de nascimento de pai, mãe e filho	controla por efeitos fixos estaduais não-observados
dummies de coorte de nascimento de pai e mãe	controla por efeitos fixos anuais

Tabela 1

### 3.2.Os Instrumentos

Utilizaremos dois grupos de instrumentos: o primeiro são séries que devem representar variações exógenas, ao nível individual, da oferta de educacional com que os pais se depararam quando estavam tomando suas decisões de escolaridade. O segundo é uma mudança institucional, a lei 5.692 de 1971, que aumentou a escolaridade mínima obrigatória de quatro para oito anos em todo o país.

#### 3.2.1. Oferta de Escolas e Professores por Unidade da Federação

Para cada adulto teremos seis instrumentos: o número de escolas existentes no estado em que ele nasceu quando tinha 7, 11 e 15 anos, e o número de professores por escola nesse estado, novamente quando o pai ou a mãe tinham 7, 11 e 15 anos. Durante o século passado, essas idades corresponderam àquelas em que a criança deveria entrar nos Ensinos Primário, Ginásio e 2º grau, respectivamente.

Esse grupo de instrumentos já foi utilizado por Emerson e Souza (2007) e Machado e Gonzaga (2007). A hipótese comum é de que tanto a quantidade de escolas por habitante quanto a de professores por escola devem estar afetando o custo e o benefício de freqüentar a escola. Além disso, não devem estar correlacionados com as características não observáveis de cada família. Nesse sentido, devem ser uma *proxy* para a oferta exógena, ao nível individual ou familiar, de educação.

#### 3.2.2. A Lei 5.692 de 1971

A reforma do ensino de 1971 incorporou à escola primária básica de quatro anos o antigo ginásial, ampliando de quatro para oito anos a escolaridade mínima obrigatória para todos os brasileiros a partir de então. Desta forma, a reforma tornou obrigatória a presença na escola para todas as crianças entre 7 e 14 anos. Foram criados o ensino de 1º grau (1ª. a 8ª série) e o ensino de 2º. grau, formado por mais três ou quatro séries anuais. Essa mudança na legislação pode ser considerada exógena ao nível familiar, e deve ter tido algum impacto mensurável sobre a escolaridade daqueles que se educavam à época. Portanto, é potencialmente um instrumento para a escolaridade dos pais.

Os pais afetados por essa reforma são os que tinham menos de 14 anos de idade em 1972, quando a lei entrou em exercício. Criamos então uma variável *dummy* que assume valor igual a 1 para todos os pais e mães nascidos antes de 1957, e zero, caso contrário.

## 4 Resultados das Estimações por MQO

### 4.1. Resultados para a Amostra Inteira

A Tabela 2 mostra o resultado das estimações de 32 especificações. As relações estimadas são: mãe-filha, mãe-filho, pai-filha e pai-filho. Na parte superior, a equação estimada segue o modelo básico. A variável dependente já definida é projetada no número de anos de estudo de um dos pais e em uma série de controles (idade do filho, idade do pai, *dummy* indicativa do sexo do filho e *dummies* estaduais). Nas especificações seguintes, são adicionadas aos controles: (i) uma variável que mede a renda per capita familiar, (ii) variáveis de características domiciliares que podem representar com maior precisão a riqueza ou “renda permanente” da família (ou do domicílio em que a criança mora), e, por último, variáveis que indicam o nível de instrução adquirido pelos avós, tanto paternos quanto maternos.

Alguns padrões interessantes surgem da análise dessas regressões simples. Primeiramente, o efeito da educação da mãe nos filhos é sempre maior que o efeito da educação do pai. Esse resultado é esperado, uma vez que a mãe pode ser considerada mais importante no processo de transmissão de capital humano que o pai, pois via de regra passa mais tempo com os filhos, especialmente nos períodos mais tenros da infância. Para todas as relações estimadas, há forte evidência de não-linearidade no processo de transmissão de capital humano: o coeficiente da variável “anos de estudo ao quadrado” é sempre altamente significativo e relativamente estável, variando de 0,12 a 0,24. Isso implica uma relação côncava entre as educações de pais e filhos: quanto menos educados, maior o efeito marginal de um ano a mais de estudo dos pais no desempenho dos filhos.

O comportamento dos coeficientes estimados nas diferentes especificações também é informativo. À medida que acrescentamos os sucessivos controles, os coeficientes de todas as relações caem bastante. Primeiro, ao incluirmos a renda per capita familiar entre os regressores, os coeficientes caem cerca de 30% para a mãe e 38% para o pai. Após os controles de variáveis referentes a características dos domicílios, os coeficientes caem ainda mais: são 46%, 41%, 56% e 51% menores, respectivamente, para as relações mãe-filha, mãe-filho, pai-filha e pai-filho, que aqueles estimados sob o modelo básico. Quando incluimos as variáveis que indicam os níveis de instrução dos avós, vemos que os coeficientes apresentam ligeira queda. De certo modo, parece que os fatores familiares não-observados que estão influenciando as variáveis de educação dos avós já foram em grande parte capturados pelas variáveis incluídas anteriormente.

	MAE-FILHA		MAE-FILHO		PAI-FILHA		PAI-FILHO		
básica	anos de estudo	<b>-2,703</b>	<b>-3,18</b>	<b>-5,296</b>	<b>-2,509</b>	<b>-4,924</b>	<b>-2,942</b>	<b>-5,308</b>	
		(0,073)***	(0,209)***	(0,065)***	(0,186)***	(0,069)***	(0,196)***	(0,065)***	
	anos de estudo <sup>2</sup>	<b>0,222</b>	<b>0,174</b>	<b>0,174</b>	<b>0,196</b>	<b>0,196</b>	<b>0,189</b>	<b>0,189</b>	
		(0,015)***	(0,013)***	(0,013)***	(0,013)***	(0,013)***	(0,013)***	(0,013)***	
Obs	18747	18747	22640	22640	18747	18747	22640	22640	
R-2	0,25	0,26	0,31	0,32	0,24	0,25	0,30	0,30	
renda	anos de estudo	<b>-1,808</b>	<b>-4,643</b>	<b>-2,261</b>	<b>-4,632</b>	<b>-1,533</b>	<b>-4,043</b>	<b>-1,918</b>	<b>-4,366</b>
		(0,085)***	(0,211)***	(0,078)***	(0,191)***	(0,081)***	(0,201)***	(0,078)***	(0,189)***
	anos de estudo <sup>2</sup>	<b>0,242</b>	<b>0,198</b>	<b>0,198</b>	<b>0,198</b>	<b>0,206</b>	<b>0,206</b>	<b>0,197</b>	<b>0,197</b>
		(0,015)***	(0,013)***	(0,013)***	(0,013)***	(0,014)***	(0,014)***	(0,013)***	(0,013)***
Obs	17792	17792	21487	21487	17792	17792	21487	21487	
R-2	0,28	0,29	0,33	0,34	0,27	0,28	0,32	0,33	
riqueza	anos de estudo	<b>-1,482</b>	<b>-3,543</b>	<b>-1,941</b>	<b>-3,695</b>	<b>-1,157</b>	<b>-2,728</b>	<b>-1,546</b>	<b>-3,21</b>
		(0,086)***	(0,21)***	(0,086)***	(0,197)***	(0,083)***	(0,202)***	(0,088)***	(0,2)***
	anos de estudo <sup>2</sup>	<b>0,177</b>	<b>0,147</b>	<b>0,147</b>	<b>0,147</b>	<b>0,13</b>	<b>0,13</b>	<b>0,134</b>	<b>0,134</b>
		(0,015)***	(0,013)***	(0,013)***	(0,013)***	(0,014)***	(0,014)***	(0,013)***	(0,013)***
Obs	17773	17773	21459	21459	17773	17773	21459	21459	
R-2	0,32	0,33	0,37	0,37	0,32	0,32	0,35	0,36	
completa	anos de estudo	<b>-1,468</b>	<b>-3,419</b>	<b>-1,883</b>	<b>-3,52</b>	<b>-1,107</b>	<b>-2,549</b>	<b>-1,454</b>	<b>-3,042</b>
		(0,091)***	(0,216)***	(0,088)***	(0,201)***	(0,09)***	(0,211)***	(0,093)***	(0,206)***
	anos de estudo <sup>2</sup>	<b>0,171</b>	<b>0,141</b>	<b>0,141</b>	<b>0,141</b>	<b>0,122</b>	<b>0,122</b>	<b>0,132</b>	<b>0,132</b>
		(0,015)***	(0,014)***	(0,014)***	(0,014)***	(0,014)***	(0,014)***	(0,014)***	(0,014)***
Obs	17773	17773	21459	21459	17773	17773	21459	21459	
R-2	0,33	0,33	0,37	0,37	0,32	0,32	0,36	0,36	

fonte: PNAD96

observações: (1) erros-padrões clusterizados por família entre parênteses; (2) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (3) var. dependente: defasagem idade-série normalizada; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, idade e idade ao quadrado de mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 2

Outro padrão interessante a ser observado é de que os coeficientes de ajustamento das regressões de filhos são sempre maiores que os das filhas. De alguma forma, o desempenho escolar das meninas parece ser mais independente das variáveis explicativas que o dos meninos.

A Tabela 3 é análoga à anterior, reportando os resultados das mesmas regressões para os dados da PNAD 88. As estimativas pontuais são consideravelmente maiores que as de 96. Os desvios-padrão estimados também são maiores, mas não na mesma proporção que os coeficientes, o que nos sugere ter havido uma queda da persistência intergeracional de capital humano entre essas datas.

De uma forma geral, as observações expostas anteriormente são válidas também para 88. O efeito da educação materna é maior que o da paterna, os efeitos sobre filhos são maiores que sobre filhas, e as meninas parecem ser mais independentes que os meninos, indicado pelo R-2 menor em todas as comparações entre os gêneros. A presença de não-linearidade nas relações também é confirmada, e para os dados de 1988 a relação parece ser ainda mais côncava que para 1996.

As variações nos coeficientes estimados à medida que incluímos os controles também seguem o padrão similar àquele dos dados de 1996, apesar de menores. Uma diferença é que as variáveis domiciliares parecem não ter relevância, enquanto a inclusão da escolaridade dos avós agora influencia bastante os coeficientes de escolaridade dos pais. A provável explicação é que as variáveis domiciliares de 1988 utilizadas no trabalho contêm menos informação que as de 1996, e por isso a inclusão de variáveis de escolaridade dos avós contribui de forma significativa para isolar o efeito da renda permanente familiar.

		MÃE-FILHA		MÃE-FILHO		PAI-FILHA		PAI-FILHO	
básica	anos de estudo	<b>-3,366</b>	<b>-5,779</b>	<b>-3,639</b>	<b>-6,16</b>	<b>-3,191</b>	<b>-5,587</b>	<b>-3,61</b>	<b>-6,457</b>
		(0,099)***	(0,315)***	(0,092)***	(0,302)***	(0,096)***	(0,31)***	(0,091)***	(0,296)***
	anos de estudo^2		<b>0,255</b>		<b>0,261</b>		<b>0,246</b>		<b>0,292</b>
			(0,03)***		(0,028)***		(0,029)***		(0,027)***
	Obs	15326	15326	18538	18538	15326	15326	18538	18538
	R-2	0,21	0,22	0,26	0,26	0,20	0,20	0,25	0,26
renda	anos de estudo	<b>-2,447</b>	<b>-5,069</b>	<b>-2,482</b>	<b>-5,138</b>	<b>-2,159</b>	<b>-4,707</b>	<b>-2,331</b>	<b>-5,275</b>
		(0,105)***	(0,311)***	(0,099)***	(0,294)***	(0,107)***	(0,311)***	(0,101)***	(0,292)***
	anos de estudo^2		<b>0,278</b>		<b>0,275</b>		<b>0,264</b>		<b>0,303</b>
			(0,029)***		(0,027)***		(0,029)***		(0,026)***
	Obs	15070	15070	18165	18165	15070	15070	18165	18165
	R-2	0,25	0,25	0,31	0,31	0,24	0,24	0,30	0,31
riqueza	anos de estudo	<b>-2,449</b>	<b>-5,061</b>	<b>-2,484</b>	<b>-5,136</b>	<b>-2,16</b>	<b>-4,695</b>	<b>-2,34</b>	<b>-5,286</b>
		(0,105)***	(0,311)***	(0,1)***	(0,294)***	(0,107)***	(0,311)***	(0,102)***	(0,292)***
	anos de estudo^2		<b>0,277</b>		<b>0,275</b>		<b>0,262</b>		<b>0,303</b>
			(0,029)***		(0,027)***		(0,029)***		(0,026)***
	Obs	15070	15070	18165	18165	15070	15070	18165	18165
	R-2	0,25	0,25	0,31	0,31	0,24	0,24	0,30	0,31
completa	anos de estudo	<b>-2,285</b>	<b>-4,737</b>	<b>-2,255</b>	<b>-4,775</b>	<b>-1,981</b>	<b>-4,489</b>	<b>-2,073</b>	<b>-4,881</b>
		(0,114)***	(0,323)***	(0,108)***	(0,3)***	(0,116)***	(0,323)***	(0,109)***	(0,3)***
	anos de estudo^2		<b>0,265</b>		<b>0,265</b>		<b>0,265</b>		<b>0,294</b>
			(0,031)***		(0,027)***		(0,03)***		(0,027)***
	Obs	15070	15070	18165	18165	15070	15070	18165	18165
	R-2	0,25	0,26	0,31	0,32	0,24	0,25	0,31	0,31

fonte: PNAD88

observações: (1) erros-padrões clusterizados por família entre parênteses; (2) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (3) var. dependente: defasagem idade-série normalizada; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, idade e idade ao quadrado de mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 3

De uma forma geral, a mobilidade intergeracional aumenta de 1988 para 1996. Talvez o processo de universalização do ensino tenha contribuído para esse resultado, mas uma análise mais cuidadosa ainda está para ser feita.

Como nessas regressões constam somente mãe ou pai, a existência de matching entre esposos e esposas pode estar viesando os resultados acima. De fato, de acordo com a equação (2), ao fazermos a projeção de  $ed_{filho}^i$  em  $ed_{mãe}^i$ , o limite em probabilidade de  $\alpha_{j,MQO}$  assume o valor  $\alpha_j + \beta_j * Cov(ed_{mãe}, ed_{pai}) / Var(ed_{pai})$ . Se  $Cov(u_{mãe}, u_{pai}) > 0$ , então a não-inclusão de educação do pai entre as variáveis explicativas introduz um viés positivo no coeficiente estimado. Por isso, a análise mais adequada nos sugere a inclusão das variáveis educacionais de ambos os pais nas especificações, para que o efeito marginal de cada um possa ser corretamente estimado.

A Tabela 4 mostra os coeficientes estimados nas especificações que incluem ambos os pais, para a PNAD96. Agora a diferença entre os efeitos materno e paterno se torna mais clara e facilmente



caracterizável: mães influenciam mais no rendimento escolar dos filhos que os pais. Na especificação completa, o efeito linear da educação da mãe sobre o filho é o dobro do efeito do pai, e sobre a filha, 78% maior. Consequentemente, a relação mais forte observada é a mãe-filho, e a mais fraca é a pai-filha, que cai a menos da metade com a inclusão dos controles de renda e de riqueza.

#### 4.2. Resultados Desagregados por Regiões

Se a mobilidade educacional depende de fatores não inclusos nas especificações estimadas, os coeficientes estimados podem variar entre as regiões na medida em que esses fatores não-observados variam entre elas. A Tabela 5 mostra as mesmas equações anteriores, separando a amostra para as regiões do Brasil. De fato, podemos notar que enquanto algumas observações anteriores se mantêm, como a maior influência materna, a magnitude dos coeficientes varia bastante entre as regiões. A região com menor mobilidade educacional é o Nordeste, enquanto Sudeste, Sul e Centro-Oeste e Norte apresentam os maiores graus de mobilidade.

	PAIS-FILHOS		PAIS-FILHA		PAIS-FILHO		
básica	anos de estudo da mãe	<b>-2,039</b> (0,068)***	<b>-3,939</b> (0,163)***	<b>-1,84</b> (0,092)***	<b>-4,026</b> (0,225)***	<b>-2,191</b> (0,086)***	<b>-3,861</b> (0,199)***
	anos de estudo do pai	<b>-1,531</b> (0,065)***	<b>-3,525</b> (0,156)***	<b>-1,397</b> (0,088)***	<b>-3,316</b> (0,211)***	<b>-1,633</b> (0,085)***	<b>-3,711</b> (0,194)***
	anos de estudo da mãe <sup>2</sup>		<b>0,156</b> (0,011)***		<b>0,182</b> (0,016)***		<b>0,135</b> (0,014)***
	anos de estudo do pai <sup>2</sup>		<b>0,157</b> (0,011)***		<b>0,15</b> (0,015)***		<b>0,163</b> (0,014)***
	Obs	41387	41387	18747	18747	22640	22640
	R-2	0,31	0,32	0,27	0,29	0,33	0,34
renda	anos de estudo da mãe	<b>-1,594</b> (0,071)***	<b>-3,589</b> (0,164)***	<b>-1,383</b> (0,096)***	<b>-3,615</b> (0,226)***	<b>-1,754</b> (0,091)***	<b>-3,563</b> (0,204)***
	anos de estudo do pai	<b>-0,982</b> (0,069)***	<b>-2,914</b> (0,158)***	<b>-0,853</b> (0,093)***	<b>-2,74</b> (0,214)***	<b>-1,083</b> (0,090)***	<b>-3,077</b> (0,199)***
	anos de estudo da mãe <sup>2</sup>		<b>0,166</b> (0,011)***		<b>0,188</b> (0,016)***		<b>0,148</b> (0,014)***
	anos de estudo do pai <sup>2</sup>		<b>0,152</b> (0,011)***		<b>0,149</b> (0,015)***		<b>0,156</b> (0,014)***
	Obs	39279	39279	17792	17792	21487	21487
	R-2	0,32	0,34	0,29	0,3	0,34	0,36
riqueza	anos de estudo da mãe	<b>-1,423</b> (0,071)***	<b>-3,005</b> (0,162)***	<b>-1,207</b> (0,095)***	<b>-2,961</b> (0,221)***	<b>-1,592</b> (0,092)***	<b>-3,039</b> (0,204)***
	anos de estudo do pai	<b>-0,77</b> (0,07)***	<b>-2,014</b> (0,159)***	<b>-0,634</b> (0,092)***	<b>-1,778</b> (0,213)***	<b>-0,875</b> (0,092)***	<b>-2,226</b> (0,203)***
	anos de estudo da mãe <sup>2</sup>		<b>0,133</b> (0,011)***		<b>0,15</b> (0,016)***		<b>0,12</b> (0,014)***
	anos de estudo do pai <sup>2</sup>		<b>0,098</b> (0,011)***		<b>0,089</b> (0,015)***		<b>0,106</b> (0,014)***
	Obs	39232	39232	17773	17773	21459	21459
	R-2	0,36	0,36	0,33	0,34	0,37	0,38
completa	anos de estudo da mãe	<b>-1,436</b> (0,072)***	<b>-2,933</b> (0,164)***	<b>-1,234</b> (0,0977)***	<b>-2,922</b> (0,225)***	<b>-1,6</b> (0,094)***	<b>-2,953</b> (0,206)***
	anos de estudo do pai	<b>-0,792</b> (0,074)***	<b>-1,951</b> (0,163)***	<b>-0,66</b> (0,0977)***	<b>-1,701</b> (0,219)***	<b>-0,9</b> (0,097)***	<b>-2,18</b> (0,208)***
	anos de estudo da mãe <sup>2</sup>		<b>0,129</b> (0,011)***		<b>0,147</b> (0,016)***		<b>0,116</b> (0,014)***
	anos de estudo do pai <sup>2</sup>		<b>0,095</b> (0,011)***		<b>0,086</b> (0,015)***		<b>0,105</b> (0,015)***
	Obs	39232	39232	17773	17773	22866	21459
	R-2	0,36	0,36	0,33	0,34	0,53	0,38

fonte: PNAD96

observações: (1) erros-padrões clusterizados por família entre parênteses; (2) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (3) var. dependente: defasagem idade-série normalizada; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, idade e idade ao quadrado de mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 4

		<i>SE</i>	<i>NE</i>	<i>SU</i>	<i>CO e NO</i>
básica	anos de estudo da mãe	<b>-1,362</b>	<b>-2,883</b>	<b>-1,788</b>	<b>-1,877</b>
		(0,106)***	(0,124)***	(0,172)***	(0,132)***
	anos de estudo do pai	<b>-1,145</b>	<b>-2,337</b>	<b>-0,96</b>	<b>-1,121</b>
		(0,100)***	(0,124)***	(0,152)***	(0,134)***
	Obs	12430	15748	5562	7647
	R-2	0,16	0,27	0,14	0,18
renda	anos de estudo da mãe	<b>-1,062</b>	<b>-2,386</b>	<b>-1,36</b>	<b>-1,539</b>
		(0,108)***	(0,132)***	(0,177)***	(0,139)***
	anos de estudo do pai	<b>-0,752</b>	<b>-1,638</b>	<b>-0,476</b>	<b>-0,832</b>
		(0,106)***	(0,135)***	(0,161)***	(0,139)***
	Obs	11953	14665	5362	7299
	R-2	0,18	0,29	0,15	0,19
riqueza	anos de estudo da mãe	<b>-0,993</b>	<b>-1,997</b>	<b>-1,29</b>	<b>-1,329</b>
		(0,111)***	(0,132)***	(0,182)***	(0,139)***
	anos de estudo do pai	<b>-0,631</b>	<b>-1,194</b>	<b>-0,36</b>	<b>-0,624</b>
		(0,111)***	(0,137)***	(0,166)**	(0,139)***
	Obs	11936	14654	5358	7284
	R-2	0,19	0,33	0,17	0,21
completa	anos de estudo da mãe	<b>-1,039</b>	<b>-2,008</b>	<b>-1,274</b>	<b>-1,287</b>
		(0,114)***	(0,136)***	(0,183)***	(0,145)***
	anos de estudo do pai	<b>-0,665</b>	<b>-1,206</b>	<b>-0,249</b>	<b>-0,659</b>
		(0,116)***	(0,144)***	(-0,176)	(0,146)***
	Obs	11936	14654	5358	7284
	R-2	0,20	0,34	0,18	0,22

fonte: PNAD96

observações: (1) erros-padrões clusterizados por família entre parênteses; (2) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (3) var. dependente: defasagem idade-série normalizada; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, idade e idade ao quadrado de mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 5

A Tabela 6 compara os coeficientes estimados para duas regiões escolhidas, o Sudeste e o Nordeste, separando por gênero da criança. Pode-se notar que os resultados mostrados na tabela acima não estão sendo gerados por um gênero em especial, mas são comuns a meninos e meninas.

		<i>SE-FILHA</i>	<i>SE-FILHO</i>	<i>NE-FILHA</i>	<i>NE-FILHO</i>
básica	anos de estudo da mãe	<b>-1,187</b>	<b>-1,542</b>	<b>-2,757</b>	<b>-2,968</b>
		(0,141)***	(0,139)***	(0,171)***	(0,154)***
	anos de estudo do pai	<b>-0,979</b>	<b>-1,252</b>	<b>-2,074</b>	<b>-2,558</b>
		(0,133)***	(0,135)***	(0,167)***	(0,154)***
	Obs	5621	6809	7260	8488
	R-2	0,14	0,17	0,24	0,28
completa	anos de estudo da mãe	<b>-0,904</b>	<b>-1,198</b>	<b>-1,82</b>	<b>-2,16</b>
		(0,147)***	(0,153)***	(0,188)***	(0,17)***
	anos de estudo do pai	<b>-0,52</b>	<b>-0,736</b>	<b>-0,937</b>	<b>-1,428</b>
		(0,147)***	(0,158)***	(0,194)***	(0,181)***
	Obs	5371	6565	6769	7885
	R-2	0,19	0,21	0,31	0,34

fonte: PNAD96

observações: (1) erros-padrões clusterizados por família entre parênteses; (2) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (3) var. dependente: defasagem idade-série normalizada; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, idade e idade ao quadrado de mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 6

As razões para esses diferentes graus de mobilidade ainda não são claras, mas a literatura sugere que aspectos associados a um menor desenvolvimento sócio-econômico, como restrições ao crédito, podem estar relacionados ao fenômeno observado. A menor oferta de escolas públicas na região Nordeste em relação à Sudeste também deve contribuir para que a mobilidade seja menor.

Na tentativa de buscar esses fundamentos, realizamos uma separação alternativa da amostra, baseada nos níveis de renda familiar.

### 4.3. Resultados Desagregados por Faixas de Renda

Uma primeira investigação sobre a possível interação entre renda e educação dos pais nos inclina a aceitar tal hipótese. Os coeficientes de variáveis construídas através da multiplicação da renda familiar por educação do pai e da mãe são ambos positivos e significantes a 1%. Para a mais adequada interpretação dos resultados, separamos as amostras em quintis de renda.

Quintis	1o	2o	3o	4o	5o
<b>PAIS-FILHOS</b>					
anos de estudo da mãe	<b>-2,690</b> (0,178)***	<b>-2,016</b> (0,171)***	<b>-1,318</b> (0,163)***	<b>-0,830</b> (0,133)***	<b>-0,476</b> (0,132)***
anos de estudo do pai	<b>-1,503</b> (0,199)***	<b>-0,918</b> (0,174)***	<b>-1,071</b> (0,172)***	<b>-0,528</b> (0,135)***	<b>-0,323</b> (0,1488)**
Obs	11341	9436	7710	6461	4284
R-2	0,29	0,25	0,19	0,18	0,13
<b>PAIS-FILHO</b>					
anos de estudo da mãe	<b>-2,596</b> (0,232)***	<b>-2,301</b> (0,209)***	<b>-1,541</b> (0,212)***	<b>-1,098</b> (0,173)***	<b>-0,575</b> (0,198)***
anos de estudo do pai	<b>-1,551</b> (0,267)***	<b>-0,986</b> (0,219)***	<b>-1,301</b> (0,217)***	<b>-0,537</b> (0,179)***	<b>-0,529</b> (0,222)**
Obs	6014	5072	4332	3589	2452
R-2	0,3	0,25	0,21	0,21	0,18
<b>PAIS-FILHA</b>					
anos de estudo da mãe	<b>-2,754</b> (0,242)***	<b>-1,611</b> (0,234)***	<b>-1,005</b> (0,217)***	<b>-0,616</b> (0,194)***	<b>-0,335</b> (0,163)**
anos de estudo do pai	<b>-1,405</b> (0,253)***	<b>-0,853</b> (0,233)***	<b>-0,758</b> (0,225)***	<b>-0,462</b> (0,198)**	<b>-0,099</b> (0,178)
Obs	5327	4364	3378	2872	1832
R-2	0,27	0,25	0,17	0,16	0,14

fonte: PNAD96  
observações: (1) erros-padrões clusterizados por família; (2) estatística-t entre parênteses; (3) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (4) var. dependente: defasagem idade-série; (5) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, idade e idade ao quadrado de mãe e pai, dummies de coorte de nascimento de mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 7

A amostra foi subdividida em cinco quintis de renda familiar per capita, e novas equações foram estimadas para cada sub-amostra. A Tabela 7 mostra os resultados para os dados da PNAD96.

Para o Brasil inteiro, vê-se de modo consistente que a influência da educação dos pais sobre os filhos decresce com o nível de renda. Em especial, o 1º quintil apresenta quase sempre as relações mais fortes, sempre altamente significantes, enquanto o 5º quintil costuma apresentar as relações mais fracas, muitas vezes perdendo a significância.

O resultado já encontrado de que os meninos são mais influenciados pela educação de pai e mãe resiste à separação da amostra em faixas de renda, com exceção dos 20% mais pobres da população. Parece emergir o seguinte padrão: à medida que aumenta a renda familiar, as filhas parecem se tornar mais “independentes” dos pais que os filhos. A isso vem se somar a maior influência da educação materna que da paterna. Como resultado, a relação pai-filha para a faixa de renda mais alta é insignificante. Caminhando em direção a níveis mais elevados de renda, os efeitos vão diminuindo, especialmente para as meninas do último quintil, sobre as quais a influência dos pais parece pouco importante, para qualquer região.

## 5 Resultados das Estimações com Variáveis Instrumentais

### 5.1. Replicando os Resultados Encontrados na Literatura

Como um primeiro exercício, reestimamos a relação que Machado e Gonzaga descrevem, mas sem utilizar a diferença entre anos de estudo de pais e avós como instrumento, uma vez que em nosso modelo essa diferença poderá estar correlacionada com fatores presentes no erro da equação de educação do filho. Para produzir resultados comparáveis, nossa medida de defasagem será transformada numa variável binária, assumindo zero se a defasagem for igual a zero e um se for maior. Também para validar a comparação, utilizaremos como controle três variáveis similares àquelas que eles constroem: a diferença entre idade da mãe e do pai, o número de filhos homens e o número de filhos mulheres. A instrumentalização da educação deve retirar os efeitos indiretos dos atributos não observados sobre o desempenho escolar do filho e nos indicar o real efeito causal existente entre essas variáveis. Espera-se, portanto, que os coeficientes estimados por 2SLS-IV sejam menores que aqueles estimados por MQO.

Os resultados que encontramos no 1º estágio - projeção de anos de estudo da mãe, do pai e renda familiar *per capita* nos instrumentos - são qualitativamente idênticos ao do trabalho citado, com a diferença de que nossos F-parciais estimados (cerca de 26, 55 e 55, respectivamente) são bem menores que em Machado e Gonzaga (cerca de 4.000, 4000 e 400), embora ainda altamente significantes. De fato, parece que as diferenças de escolaridades entre pais e avós são responsáveis por parte do alto poder explicativo daqueles instrumentos para as escolaridades de mãe e pai. Ainda assim, os resultados do 2º

estágio dos dois trabalhos, como mostrado na Tabela 8, são qualitativamente idênticos. Além disso, as estimativas pontuais de educação dos pais sob esse método são menores que as estimativas por MQO, resultado condizente com o viés esperado pelo modelo de fatores familiares não-observáveis. Nossa estimativa também é similar àquela de Oreopoulos *et al.* (2006), que encontram um efeito entre 2% e 4%.

Efeitos estimados	este trabalho (2008)	Machado e Gonzaga (2007)
anos de estudo da mãe	<b>-0.032</b> (0.016)**	<b>-0.011</b> (0.003)***
anos de estudo do pai	<b>-0.005</b> (0.013)	<b>-0.003</b> (0.004)
log da renda familiar per capita	<b>-0.17</b> (0.020)***	<b>-0.168</b> (0.049)***
Obs	40332	20332
R-2	0.24	0.26

fonte: PNAD96, IBGE e Machado e Gonzaga (2008)  
 observações para a especificação: (1) variável dependente: defasagem idade-série binária; (2) erros-padrões clusterizados por família entre parênteses; (3) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós, diferença entre idade do pai e da mãe, número de irmãos, número de irmãs.

Tabela 8

## 5.2. Especificação Básica

Passamos agora a especificações mais próximas da literatura internacional. A variável dependente é aquela do capítulo anterior. Inicialmente, retiramos da especificação “completa” do capítulo 4 as idades dos pais. A especificação básica desse capítulo, portanto, possui controles de (i) idade e idade ao quadrado dos filhos, (ii) gênero e raça, (iii) renda familiar per capita, (iv) variáveis domiciliares (*proxies* de riqueza), (v) nível educacional atingido pelos avós maternos e paternos (*proxies* de riqueza e *background* familiar) e (vi) efeitos fixos estaduais para os estados natais de filhos, mãe e pai.

### 5.2.1. Efeito da Mãe

Assim como no capítulo anterior, iremos estimar efeitos de pai e mãe separadamente, a princípio. A Tabela 9 mostra a projeção da educação da mãe nas variáveis escolhidas como seus instrumentos, para a amostra com todos os filhos.

Tanto para a PNAD96 quanto para a PNAD88 os testes F-Parciais de significância conjunta dos instrumentos indicam forte correlação parcial dessas variáveis com a educação da mãe. É interessante notar que há significância individual de variáveis de ambos os grupos que representam a oferta educacional: escolas por habitante e professores por escola. A Tabela 10 mostra, para os dados da PNAD96, a matriz de correlações das variáveis utilizadas como instrumentos. As correlações negativas entre as variáveis pertencentes aos dois grupos sugerem que eles parecem representar dimensões diferentes da oferta educacional em cada estado. Nesse sentido, parece haver duas formas distintas de um estado melhorar sua oferta educacional: construindo mais escolas ou contratando mais professores por escola. Em especial, no 1º estágio vemos que para ambas as PNADs são significantes a 1% (e no sentido esperado) as variáveis (i) número de escolas por habitante quando a mãe tinha 7 anos; (ii) escolas por habitante quando ela tinha 15 anos; e (iii) professores por escola quando ela tinha 15 anos. Lembrando que os efeitos fixos estaduais já estão controlados, parece haver evidência de validade dos instrumentos.

Quantitativamente, os efeitos são importantes: com *backgrounds* familiares exatamente iguais, caso uma mãe tenha se educado quatro anos a mais que outra espera-se que seu filho tenha uma defasagem idade-série *normalizada* cerca de 5 p.p. menor. Todos os coeficientes são significantes a 1%, com exceção daqueles sobre a filha para os dados de 1988, que é significativa a 5%, e para os 1996, que não se mostra significativa. Assim como na análise por MQO e como encontrado por Black *et al.* (2005), os efeitos são notadamente maiores para meninos que para meninas.

*var dependente: anos de estudos da mãe*

	PNAD96	PNAD88
esc7_mae	0.437 (3.11)***	0.501 (3.13)***
esc11_mae	0.158 (0.90)	0.054 (0.26)
esc15_mae	0.442 (3.40)***	0.577 (3.46)***
prof7_mae	-0.005 (0.11)	-0.065 (1.34)
prof11_mae	0.019 (0.43)	0.109 (2.47)**
prof15_mae	0.172 (4.98)***	0.151 (4.78)***
nasc_mae_1933_1956	-0.057 (0.81)	0.075 (1.32)
Obs	41317	30986
R-2	0.47	0.36
<b>F-Parcial</b>	<b>68.36</b>	<b>62.57</b>
<b>Prob&gt;F</b>	<b>0.00</b>	<b>0.00</b>

fonte: PNAD96, PNAD88 e IBGE

observações: (1) erros-padrões clusterizados por família; (2) estatística-t entre parênteses; (3) \* p-valor<=10%.

\*\* p-valor<=5%, \*\*\* p-valor<=1%; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem,

branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de

características do domicílio e dummies de escolaridade dos avós

Tabela 9

**Correlações entre as Variáveis de Oferta Educacional da Mãe**

	esc7	esc11	esc15	prof7	prof11	prof15
esc7	1,00					
esc11	0,92	1,00				
esc15	0,81	0,92	1,00			
prof7	-0,35	-0,46	-0,57	1,00		
prof11	-0,33	-0,47	-0,57	0,97	1,00	
prof15	-0,33	-0,47	-0,59	0,95	0,97	1,00

(obs=43643)

Fonte: PNAD96 e IBGE

Tabela 10

A Tabela 11 mostra o 2º estágio da estimação. Caso a estratégia de 2SLS-IV esteja correta, os coeficientes estimados são interpretados como o efeito *causal* de um ano a mais de estudo da mãe na defasagem idade-série dos filhos.

**especificação básica**

	FILHOS		FILHO		FILHA	
PNAD	96	88	96	88	96	88
anos de estudo da mãe	<b>-1,224</b> (0,449)***	<b>-1,735</b> (0,499)***	<b>-1,574</b> (0,590)***	<b>-1,848</b> (0,612)***	<b>-0,879</b> (0,582)	<b>-1,575</b> (0,675)**
Obs	38665	30986	21157	16915	17508	14071
R-2	0,36	0,30	0,37	0,33	0,33	0,26

fonte: PNAD96, PNAD88 e IBGE

observações: (1) variável dependente: defasagem idade-série; (2) erros-padrões clusterizados por família; (3) estatística-t entre parênteses; (4) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (5) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 11

Em relação aos coeficientes estimados no capítulo anterior, os coeficientes de IVs são menores que os de MQO para os dados de 1996 e 1988, como prevê a teoria. Para a PNAD96 os coeficientes caem de -1,883 para -1,574 para os filhos, e de -1,468 para -0,879 para as filhas. Para a PNAD88 os coeficientes caem de -2,255 para -1,848 para os filhos, e de -2,285 para -1,575 para as filhas.

**5.2.2. Efeito do Pai**

Faremos o exercício análogo para o efeito da instrução do pai. A Tabela 12 mostra o 1º estágio da estimação por 2SLS do efeito causal da educação do pai sobre a defasagem idade-série do filho. O padrão das estimativas e dos F-Parciais é bastante semelhante àquele observado na seção anterior. Também vemos estimativas significantes e no sentido esperado de variáveis de ambos os grupos (escolas/hab e professores/escola). Em particular, as variáveis significantes a 1% para ambas as PNADs são (i) escolas por habitante quando o pai tinha 7 anos; (ii) professores por escola quando a pai tinha 15 anos. Portanto, tanto para pai quanto para mãe surge um padrão interessante: as variáveis mais robustamente significantes são número de escolas quando tinha 7 anos de idade e número de professores por escola quando tinham 15 anos.

*var. dependente: anos de estudo do pai*

	PNAD96	PNAD88
esc7_pai	0.517 (3.75)***	0.470 (2.66)***
esc11_pai	0.212 (1.21)	0.126 (0.59)
esc15_pai	0.151 (1.24)	0.444 (2.64)***
prof7_pai	-0.049 (1.05)	-0.018 (0.31)
prof11_pai	0.040 (0.77)	0.115 (1.90)*
prof15_pai	0.197 (5.59)***	0.119 (2.82)***
nasc_pai_1933_1956	0.028 (0.50)	0.258 (4.04)***
Obs	40584	28946
R-2	0.52	0.41
<b>F-Parcial</b>	<b>69.39</b>	<b>53.56</b>
<b>Prob&gt;F</b>	<b>0.00</b>	<b>0.00</b>

fonte: PNAD96, PNAD88 e IBGE

observações: (1) erros-padrões clusterizados por família; (2) estatística-t entre parênteses; (3) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 12

Os instrumentos para educação paterna mostram o mesmo padrão daqueles para a educação materna, como mostra abaixo a Tabela 13.

**Correlações entre as Variáveis de Oferta Educacional do Pai**

	esc7	esc11	esc15	prof7	prof11	prof15
esc7	1,00					
esc11	0,93	1,00				
esc15	0,83	0,93	1,00			
prof7	-0,24	-0,34	-0,45	1,00		
prof11	-0,20	-0,33	-0,44	0,97	1,00	
prof15	-0,18	-0,31	-0,45	0,94	0,97	1,00

(obs=42857)

Fonte: PNAD96 e IBGE

Tabela 13

A Tabela 14 mostra o 2º estágio da estimação. Para a PNAD96, a estimativa pontual do efeito causal da educação do pai sobre a filha é significativa, e a estimativa do efeito sobre o filho insignificante. Já para a PNAD88, ocorre o inverso. Comparando com os resultados estimados no capítulo anterior, vemos que para a PNAD96 a estimativa do coeficiente sobre o filho cai de -1,454 para -0,839 (erro-padrão sobe de 0,093 para 0,589), enquanto o coeficiente do efeito sobre a filha sobe de -1,107 para -1,809 (erro-padrão vai de 0,09 para 0,627). Já para os dados da PNAD88, observamos uma queda dos coeficientes em relação àqueles estimados pelo método MQO. O coeficiente da relação pai-filho cai de -2,073 para -1,968 (e seu erro-padrão sobe de 0,109 para 0,66), enquanto o coeficiente da relação pai-filha cai de -1,981 para -0,693, tornando-se insignificante (erro-padrão sobe de 0,116 para 0,76).

Os resultados diferentes entre os anos podem estar refletindo mudanças em algumas estruturas, mas devemos pesquisar mais para maior compreensão. Além disso, como demonstrado no capítulo 2, a inclusão de somente um dos pais muito provavelmente está viesando nossos resultados. A próxima seção trata de tentar separar esses efeitos.

especificação básica

	FILHOS		FILHO		FILHA	
PNAD	96	88	96	88	96	88
anos de estudo do pai	<b>-1,287</b> (0,470)***	<b>-1,349</b> (0,541)**	<b>-0,839</b> (0,589)	<b>-1,968</b> (0,660)***	<b>-1,809</b> (0,627)***	<b>-0,693</b> (0,76)
Obs	37963	28946	20794	15844	17169	13102
R-2	0,35	0,29	0,36	0,32	0,32	0,24

fonte: PNAD96, PNAD88 e IBGE

observações: (1) variável dependente: defasagem idade-série; (2) erros-padrões clusterizados por família; (3) estatística-t entre parênteses; (4) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (5) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 14

### 5.2.3.Efeitos Parciais de Mãe e Pai

Para separar os efeitos de mãe e pai, a estratégia é a mesma do capítulo anterior. No 1º estágio, projetamos anos de estudo de mãe e pai nos instrumentos, que agora são quatorze. A Tabela 15 mostra os efeitos parciais estimados dos instrumentos nas variáveis de educação da mãe e do pai. Os F-parciais dos instrumentos, todos entre 28 e 40, sugerem forte correlação parcial dos instrumentos com as variáveis supostas endógenas.

PNAD	variável dependente			
	anos de estudo da mãe		anos de estudo do pai	
	96	88	96	88
esc7_mae	0.065 (0.45)	0.390 (2.19)**	0.061 (0.46)	0.032 (0.19)
esc11_mae	0.188 (1.06)	-0.157 (0.72)	0.107 (0.65)	-0.102 (0.49)
esc15_mae	0.248 (1.82)*	0.408 (2.25)**	0.070 (0.56)	0.427 (2.62)***
prof7_mae	-0.055 (1.29)	-0.073 (1.31)	0.027 (0.65)	-0.051 (0.83)
prof11_mae	0.006 (0.14)	0.093 (1.81)*	-0.068 (1.49)	0.034 (0.65)
prof15_mae	0.129 (3.74)***	0.094 (2.66)***	0.122 (3.50)***	0.106 (2.81)***
esc7_pai	0.627 (4.13)***	0.168 (0.89)	0.403 (2.90)***	0.378 (2.14)**
esc11_pai	0.064 (0.33)	0.262 (1.16)	0.208 (1.19)	0.068 (0.32)
esc15_pai	0.075 (0.54)	0.040 (0.22)	0.003 (0.02)	0.329 (1.88)*
prof7_pai	0.040 (0.77)	0.039 (0.58)	-0.038 (0.82)	-0.018 (0.31)
prof11_pai	0.018 (0.32)	0.095 (1.38)	0.030 (0.58)	0.115 (1.78)*
prof15_pai	0.077 (2.06)**	0.049 (1.02)	0.131 (3.62)***	0.061 (1.34)
nasc_pai_1933_1956	0.185 (3.09)***	0.224 (3.41)***	0.051 (0.88)	0.164 (2.45)**
nasc_mae_1933_1956	-0.071 (0.97)	-0.042 (0.63)	-0.167 (2.38)**	0.049 (0.75)
Obs	40284	28308	40284	28308
R-2	0.48	0.37	0.53	0.44
<b>F-Parcial</b>	<b>40.00</b>	<b>29.89</b>	<b>37.88</b>	<b>28.23</b>
<b>Prob&gt;F</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>

fonte: PNAD96, PNAD88 e IBGE

observações: (1) erros-padrões clusterizados por família; (2) estatística-t entre parênteses; (3) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 15

No entanto, a análise de cada instrumento individualmente mostra resultados inesperados. Para a PNAD96, significativamente correlacionados com educação da mãe temos: (i) escolas por habitante quando a mãe tinha 15 anos; (ii) professores por escola quando ela tinha 15 anos; (iii) escolas por habitante *no estado de nascimento do pai quando ele tinha 7 anos*; (iv) professores por escola *no estado de nascimento do pai quando ele tinha 15 anos*; (v) variável *dummy* indicativa de se o pai nasceu antes de 1957. Ou seja: significativamente correlacionados com a educação da mãe temos *duas* variáveis que

esperávamos estar influenciando diretamente a educação da mãe e *três* que esperávamos estar influenciando diretamente a educação do pai. Analisando as três outras colunas, vemos que o fenômeno se repete: instrumentos que esperávamos representar variações significantes nos ambientes em que mães ou pais viveram quando estavam em sua idade escolar influenciam mãe e pai *indistintamente*.

A Tabela 16 mostra a correlação existente entre os instrumentos. A combinação de: (i) correlação entre as variáveis de oferta educacional de mãe ou pai, (ii) correlação entre os estados de nascimento de mãe e pai e (iii) correlação entre os anos de nascimento de mãe e pai; faz com que os instrumentos de mãe e pai sejam bastante correlacionados entre si. Isso explica os resultados descritos no parágrafo anterior, e dificulta a estratégia de identificação proposta. Esse problema também foi encontrado por Oreopoulos *et al.* (2006) e, provavelmente, também por Black *et al.* (2005), o que deve tê-los levado a não se concentrarem em pai e mãe simultaneamente.

Correlações entre os Instrumentos														
	esc7_mae	esc11_mae	esc15_mae	prof7_mae	prof11_mae	prof15_mae	esc7_pai	esc11_pai	esc15_pai	prof7_pai	prof11_pai	prof15_pai	nasc. mãe antes de 1957	nasc. pai antes de 1957
esc7_mae	1,00													
esc11_mae	0,92	1,00												
esc15_mae	0,81	0,92	1,00											
prof7_mae	-0,36	-0,47	-0,57	1,00										
prof11_mae	-0,34	-0,48	-0,58	0,97	1,00									
prof15_mae	-0,34	-0,48	-0,60	0,95	0,97	1,00								
esc7_pai	0,68	0,61	0,52	-0,12	-0,12	-0,12	1,00							
esc11_pai	0,69	0,67	0,60	-0,22	-0,22	-0,22	0,93	1,00						
esc15_pai	0,66	0,69	0,67	-0,31	-0,32	-0,33	0,83	0,93	1,00					
prof7_pai	-0,22	-0,31	-0,39	0,66	0,65	0,65	-0,24	-0,34	-0,46	1,00				
prof11_pai	-0,21	-0,31	-0,40	0,67	0,68	0,67	-0,20	-0,33	-0,45	0,97	1,00			
prof15_pai	-0,21	-0,31	-0,40	0,67	0,69	0,69	-0,19	-0,31	-0,45	0,94	0,97	1,00		
nasc. mãe antes de 1957	-0,36	-0,30	-0,22	-0,33	-0,31	-0,25	-0,31	-0,27	-0,22	-0,21	-0,24	-0,23	1,00	
nasc. pai antes de 1957	-0,20	-0,16	-0,08	-0,20	-0,17	-0,15	-0,31	-0,25	-0,19	-0,32	-0,31	-0,25	0,46	1,00

(obs=42534)  
Fonte: PNAD96 e IBGE

Tabela 16

A Tabela 17 mostra o resultado da estimação através do método de 2SLS. Os resultados das estimativas pontuais tornam-se confusos: alguns coeficientes assumem valor positivo, embora a maioria permaneça negativa. A única equação que produz ambas as estimativas negativas está na 1ª coluna (filhos e filhas, PNAD96). Todas as outras reportam um coeficiente negativo e outro positivo.

Soma-se a esse cenário o enorme aumento nos erros-padrão, a ponto de nenhuma das relações ser significativa, tanto para 1996 quanto para 1988. Claramente, o método não é capaz de distinguir de forma adequada entre efeitos de mãe e pai.

#### especificação básica

	FILHOS		FILHO		FILHA	
PNAD	96	88	96	88	96	88
anos de estudo da mãe	<b>-0,654</b>	<b>0,324</b>	<b>-1,562</b>	<b>1,083</b>	<b>0,013</b>	<b>-2,354</b>
	(1,473)	(2,071)	(1,632)	(2,022)	(1,735)	(2,528)
anos de estudo do pai	<b>-0,65</b>	<b>-1,96</b>	<b>0,405</b>	<b>-3,212</b>	<b>-1,45</b>	<b>1,207</b>
	(1,547)	(2,199)	(1,693)	(2,201)	(1,808)	(2,658)
Obs	37688	28308	20640	15503	17048	12805
R-2	0,35	0,29	0,36	0,30	0,32	0,24

fonte: PNAD96, PNAD88 e IBGE

observações: (1) variável dependente: defasagem idade-série; (2) erros-padrões clusterizados por família; (3) estatística-t entre parênteses; (4) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (5) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 17

A alta correlação entre educação materna e paterna ( $p=0,71$ ), combinada com a alta correlação entre os instrumentos utilizados para mãe e pai dificulta muito a identificação dos efeitos parciais do nível de instrução de cada um deles. As variações entre os instrumentos utilizados não são suficientes para separar de forma correta os aumentos de instrução de pai e mãe separadamente, e a existência de *matching* no mercado de casamento torna os resultados confusos.

#### 5.2.4.Efeito Conjunto do Capital Humano dos Pais

Uma vez que não conseguimos isolar os efeitos parciais, iremos adotar a estratégia de Oreopoulos *et al.* (2006). Na verdade, nosso ambiente é muito parecido com o de seu estudo: os instrumentos utilizados têm como fonte de variação o par ordenado estado/ano de nascimento do pai ou mãe. Como solução, construímos uma nova variável: a soma dos anos de estudo completos dos pais, buscando isolar o seu efeito causal.



Antes de partir para o método 2SLS, é informativo estimar as relações via MQO. A Tabela 18 mostra o resultado dessas estimações para a especificação que no capítulo anterior chamamos “completa”. Vemos que os principais resultados são condizentes com aquilo que já encontramos: maior mobilidade em 1996 *vis-à-vis* 1988 e menor mobilidade de meninos em relação às meninas.

especificação básica

	FILHOS		FILHO		FILHA	
PNAD	96	88	96	88	96	88
anos de estudo dos pais	-1,119 (0,042)***	-1,598 (0,051)***	-1,256 (0,054)***	-1,595 (0,065)***	-0,950 (0,055)**	-1,594 (0,069)**
Obs	39232	33235	21459	18165	17773	15070
R-2	0,36	0,30	0,37	0,32	0,33	0,26

fonte: PNAD96, PNAD88 e IBGE

observações: (1) variável dependente: defasagem idade-série; (2) erros-padrões clusterizados por família; (3) estatística-t entre parênteses; (4) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (5) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 18

Partimos agora para a estimação via variáveis instrumentais. A Tabela 19 mostra o resultado do 1º estágio. Como esperado, as estatísticas-F indicam forte correlação parcial dos instrumentos com a nova variável endógena. Para ambas as PNADs há variáveis de oferta de escola e de professores por escola que se mostram significantes, reforçando a argumentação de que constituem dimensões distintas da oferta educacional.

var. dependente: anos de estudo dos pais

PNAD	96	88
esc7_mae	0,127 (0,55)	0,422 (1,51)
esc11_mae	0,295 (1,06)	-0,259 (0,76)
esc15_mae	0,318 (1,5)	0,835 (2,97)***
prof7_mae	-0,028 (0,42)	-0,124 (1,29)
prof11_mae	-0,062 (0,86)	0,127 (1,51)
prof15_mae	0,251 (4,49)***	0,2 (3,39)***
esc7_pai	1,030 (4,35)***	0,546 (1,84)*
esc11_pai	0,272 (0,91)	0,33 (0,9)
esc15_pai	0,078 (0,36)	0,369 (1,26)
prof7_pai	0,002 (0,03)	0,02 (0,19)
prof11_pai	0,048 (0,56)	0,21 (1,96)*
prof15_pai	0,208 (3,56)***	0,111 -1,48
nasc_pai_1933_1956	0,236 (2,47)**	0,388 (3,71)***
nasc_mae_1933_1956	-0,238 (2,03)**	0,007 (0,06)
Obs	40284	28308
R-2	0,6	0,5
<b>F-Parcial</b>	56,6	43,65
<b>Prob&gt;F</b>	0,00	0,00

fonte: PNAD96, PNAD88 e IBGE

observações: (1) erros-padrões clusterizados por família; (2) estatística-t entre parênteses; (3) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (4) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 19

A Tabela 20 mostra o resultado do 2º estágio da estimação através do método 2SLS. Os coeficientes estimados sugerem que, realmente, existe efeito causal de anos de estudo dos pais no desempenho escolar dos filhos. Todas as relações são significantes a 1%, com exceção de pais-filha para a PNAD88, que é significativa a 10%. As estimativas pontuais são todas menores que aquelas estimadas via MQO, indicando a correção do viés de endogeneidade causado por fatores familiares não-observáveis. Não é

claro se há diferença entre os anos ou o gênero dos filhos, à parte um maior desvio-padrão das estimativas para a PNAD88.

especificação básica

	FILHOS		FILHO		FILHA	
PNAD	96	88	96	88	96	88
anos de estudo dos pais	<b>-0,652</b>	<b>-0,776</b>	<b>-0,605</b>	<b>-0,967</b>	<b>-0,700</b>	<b>-0,622</b>
	(0,212)***	(0,249)***	(0,268)**	(0,309)***	(0,286)**	(0,341)*
Obs	37688	28308	20640	15503	17048	12805
R-2	0,35	0,30	0,37	0,33	0,33	0,26

fonte: PNAD96, PNAD88 e IBGE

observações: (1) variável dependente: defasagem idade-série; (2) erros-padrões clusterizados por família; (3) estatística-t entre parênteses; (4) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (5) controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 20

A magnitude dos efeitos também é relevante: um ano a mais de estudo de pai *ou* mãe implica uma diminuição de 0,7 pontos na nossa medida de atraso escolar. Isso sugere que, para duas famílias com idênticas variáveis explicativas, um aumento de 4 anos de estudo para pai *e* mãe tem um efeito direto de diminuir a medida de atraso escolar em 5,6 pontos. Para o caso de um(a) filho(a) no fim do ciclo escolar, com 17 anos, espera-se um atraso de 0,56 ano menor.

Para melhor analisar os mecanismos por trás das correlações entre educação observada dos pais e seus instrumentos, partimos para outras especificações, incluindo controles de idade dos pais e de coorte, utilizados em estratégia empírica mais próxima a Black *et al.* (2005) e Oreopoulos *et al.* (2006).

### 5.3.Outras especificações

Na seção anterior, mostramos que os instrumentos de mãe e pai influenciam a escolaridade de ambos de forma indistinta. É possível que não estejamos conseguindo diferenciar corretamente os efeitos dos instrumentos devido às correlações entre as variáveis e ao *matching* entre pais e mães; ou, ainda, que esses instrumentos estejam captando outros fatores que afetaram tanto a instrução do pai quanto a da mãe, além daqueles diretamente relacionados ao suporte teórico utilizado inicialmente, que era de representarem uma variação na oferta educacional com impacto sobre as decisões de escolaridade das crianças à época. Como nossos instrumentos são construídos com base em séries de escolas e professores ao longo do século XX, podemos estar capturando efeitos de outros fatores que estiveram variando ao longo do tempo de forma correlacionada com eles.

Certamente, muitos fatores ao longo do século passado podem estar afetando não só a oferta educacional como também a demanda por escolaridade. Idealmente, gostaríamos de medir esses outros fatores e encontrar o efeito de cada um deles sobre a escolaridade dos indivíduos e, posteriormente, construir uma hipótese de identificação baseada naqueles fatores que realmente poderíamos considerar bons instrumentos. Caso eles sejam independentes de fatores afetando a escolaridade dos filhos da amostra (ou controláveis), podem ser utilizados como fonte de variações exógenas necessárias a uma estratégia de 2SLS-IV.

Nesse trabalho não seremos capazes de identificar quais são esses outros fatores, mas podemos tentar isolar o efeito parcial dos nossos instrumentos se entre os regressores incluirmos uma tendência temporal ou, equivalentemente, a idade da mãe ou do pai.

A segunda coluna da Tabela 21 mostra os F-parciais dos instrumentos sob essa nova especificação. O efeito das variáveis de escolas e professores na educação dos pais ainda é significativo a 1%, porém os valores dos testes F-parciais caem consideravelmente ao acrescentarmos a idade da mãe ou do pai entre os controles. Ao incluirmos também termos quadráticos das idades, vemos que o poder preditivo dos instrumentos começa a se tornar fraco, nos sugerindo que outros fatores correlacionados com uma tendência temporal comum aos instrumentos, não necessariamente linear, parecem estar explicando os fortes valores dos F-parciais encontrados anteriormente.

Seguindo a especificação típica dos estudos de transmissão intergeracional de capital humano internacionais, tais como Black *et al.* (2005) e Oreopoulos *et al.* (2006), incluímos agora dummies de coorte para nascimento dos pais, que devem controlar eventuais efeitos fixos para os anos de nascimento de pai e de mãe. Ao fazermos isso, os valores dos F-parciais dos instrumentos se torna muito baixo, como mostra a coluna sob a especificação que chamamos “completa”.

PNAD96								
var. dependente:	anos de estudo da mãe				anos de estudo do pai			
	basica	idades	idades^2	completa	basica	idades	idades^2	completa
especificações:								
esc7_mae	0.065 (0.45)	-0.325 (1.91)*	-0.212 (1.25)	-0.198 (1.08)	0.061 (0.46)	-0.205 (1.30)	-0.140 (0.89)	-0.109 (0.63)
esc11_mae	0.188 (1.06)	0.219 (1.23)	0.266 (1.50)	0.262 (1.29)	0.107 (0.65)	0.155 (0.94)	0.191 (1.16)	0.130 (0.68)
esc15_mae	0.248 (1.82)*	0.141 (1.02)	-0.102 (0.70)	-0.102 (0.65)	0.070 (0.56)	-0.005 (0.04)	-0.139 (1.03)	-0.077 (0.52)
prof7_mae	-0.055 (1.29)	-0.105 (2.34)**	-0.034 (0.75)	0.007 (0.13)	0.027 (0.65)	-0.001 (0.03)	0.039 (0.91)	0.055 (1.13)
prof11_mae	0.006 (0.14)	0.018 (0.41)	-0.001 (0.02)	-0.022 (0.45)	-0.068 (1.49)	-0.053 (1.16)	-0.063 (1.38)	-0.065 (1.33)
prof15_mae	0.129 (3.74)***	0.049 (1.32)	0.024 (0.63)	0.010 (0.26)	0.122 (3.50)***	0.059 (1.55)	0.043 (1.12)	0.033 (0.81)
esc7_pai	0.627 (4.13)***	0.137 (0.79)	0.241 (1.37)	0.188 (1.01)	0.403 (2.90)***	-0.189 (1.19)	-0.093 (0.58)	-0.082 (0.47)
esc11_pai	0.064 (0.33)	0.046 (0.24)	0.038 (0.20)	-0.084 (0.40)	0.208 (1.19)	0.162 (0.92)	0.160 (0.91)	0.064 (0.32)
esc15_pai	0.075 (0.54)	-0.091 (0.62)	-0.124 (0.81)	-0.054 (0.33)	0.003 (0.02)	-0.233 (1.71)*	-0.296 (2.10)**	-0.356 (2.32)**
prof7_pai	0.040 (0.77)	-0.054 (0.97)	-0.009 (0.16)	-0.016 (0.23)	-0.038 (0.82)	-0.153 (3.08)***	-0.113 (2.25)**	-0.054 (0.92)
prof11_pai	0.018 (0.32)	0.010 (0.19)	-0.015 (0.27)	0.028 (0.44)	0.030 (0.58)	0.015 (0.29)	-0.004 (0.08)	-0.063 (1.07)
prof15_pai	0.077 (2.06)**	0.020 (0.52)	0.025 (0.64)	-0.022 (0.49)	0.131 (3.62)***	0.054 (1.41)	0.057 (1.49)	0.032 (0.75)
nasc_pai_1933_1956	0.185 (3.09)***	0.139 (2.30)**	-0.031 (0.42)	-0.120 (0.38)	0.051 (0.88)	-0.003 (0.05)	-0.156 (2.10)**	-0.124 (0.42)
nasc_mae_1933_1956	-0.071 (0.97)	0.065 (0.85)	-0.034 (0.44)	-0.080 (0.17)	-0.167 (2.38)**	-0.047 (0.64)	-0.103 (1.37)	-0.282 (0.58)
Obs	40284	40284	40284	40284	40284	40284	40284	40284
R-2	0.48	0.48	0.48	0.49	0.53	0.54	0.54	0.54
<b>F-Parcial</b>	<b>40.00</b>	<b>3.00</b>	<b>0.65</b>	<b>0.33</b>	<b>37.88</b>	<b>1.81</b>	<b>2.24</b>	<b>1.63</b>
<b>Prob&gt;F</b>	<b>0.00</b>	<b>0.00</b>	<b>0.82</b>	<b>0.99</b>	<b>0.00</b>	<b>0.03</b>	<b>0.00</b>	<b>0.06</b>

fonte: PNAD96 e IBGE

observações: (1) erros-padrões clusterizados por família; (2) estatística-t entre parênteses; (3) \* p-valor<=10%, \*\* p-valor<=5%, \*\*\* p-valor<=1%; (4)

controles: idade em fevereiro, idade em fevereiro ao quadrado, homem, branco, dummies de UF de nascimento de filho, mãe e pai, idade e idade ao quadrado de mãe e pai, dummies de coorte de nascimento de mãe e pai, ln da renda familiar per capita, dummies de características do domicílio e dummies de escolaridade dos avós

Tabela 21

Esses resultados nos sugerem que não há variação suficiente nas variáveis de escolas e professores além daquelas observadas ao nível de coorte e de estado que nos permita a identificação inteiramente ortogonal a esses fatores. Nossa identificação inicial, portanto, pode estar associada a outras mudanças ocorrendo ao nível da coorte, tais como aumento de renda ou provisão de outros bens públicos. Incluindo séries históricas relacionadas a outras dimensões de bens públicos entre os controles, talvez sejamos capazes de isolar de forma mais correta o efeito de variação em oferta educacional que estamos procurando.

Além disso, é bem provável que essas variáveis tenham um erro de medida particularmente grande, por isso, seus efeitos podem estar subestimados. Talvez a construção alternativa desses instrumentos dê resultados mais satisfatórios, como, por exemplo, a desagregação entre oferta de escolas públicas e privadas ou ao nível municipal, se possível.

Mas, não se pode descartar a possibilidade de simplesmente esses dados serem uma *proxy* muito imperfeita para a oferta educacional com que pai e mãe se depararam. Dentre os fatores que podem explicar o efeito que capturamos inicialmente, podemos considerar o aumento do nível de renda brasileiro durante o século passado, que pode ter aumentado a demanda por educação, o crescimento da provisão de bens públicos, ou ainda uma melhoria nas condições de saúde que também se refletiria em maior retorno intertemporal à educação adquirida. Certamente, uma pesquisa mais detalhada pode ajudar a entender as fontes de variação no nível educacional observado entre coortes.

## 6 Conclusão

Neste trabalho, analisamos o processo de transmissão intergeracional de capital humano no Brasil. Estimamos quanto um ano a mais de escolaridade dos pais influencia na defasagem idade-série de seus filhos, buscando controlar tanto para renda e riqueza, canais pelos quais essa transmissão se dá de forma indireta, quanto para fatores familiares não-observados, através da inclusão de variáveis de escolaridade dos avós. Estimando separadamente as relações para filhos e filhas, damos liberdade para que os efeitos da escolaridade de mãe e pai sejam diferentes sobre meninos e meninas. Em seguida, através do método

de Mínimos Quadrados em Dois Estágios com o uso de variáveis instrumentais, tentamos isolar os efeitos causais de escolaridade dos pais sobre a defasagem idade-série dos filhos.

As estimativas MQO mostram que há evidência de diferentes efeitos de mãe e pai sobre filho e filha. Mães afetam mais os filhos que os pais, resultado também encontrado em diversos artigos na literatura. Meninas parecem ser mais independentes da influência dos pais que meninos. A persistência intergeracional de capital humano varia consideravelmente entre as regiões do país, e é particularmente alta no Nordeste. Separando por faixas de renda, vemos que esse resultado pode estar associado ao fato de que as camadas mais pobres da população exibem mobilidade intergeracional muito menor que as mais ricas.

Utilizando séries históricas de escolas e professores por estado no século XX como instrumento, encontramos que, analisados separadamente, tanto escolaridade paterna quanto materna têm efeitos causais diretos sobre a defasagem idade-série dos filhos. Quando tentamos separar os efeitos parciais de pai e mãe, não conseguimos distingui-los de forma satisfatória. A utilização da variável construída através da soma dos anos de estudo dos pais indica um claro efeito causal de educação dos pais em desempenho escolar dos filhos, de magnitude semelhante àquela encontrada na literatura.

A análise do processo de instrumentalização utilizado mostra que as variáveis escolhidas como instrumentos estão muito correlacionadas com efeitos de coorte, podendo estar refletindo outros aspectos que variaram durante o século passado e que também influenciaram a escolaridade dos pais. Nesse sentido, é necessário um estudo mais aprofundado de quais são as fontes de variação exógenas e identificáveis que serviriam como potenciais instrumentos para educação dos pais.

## 7 Bibliografia

- [1] Becker, Gary S. e Nigel Tomes (1986). “Human Capital and the Rise and Fall of Families”. *Journal of Labor Economics*, Vol. 4, No. 3, Part 2 (Jul., 1986), pp. S1-S39.
- [2] Black, Sandra E., Paul J. Devereux, e Kjell G. Salvanes (2005). “Why the Apple Doesn’t Fall Far: Understanding the Intergenerational Transmission of Human Capital”. *American Economic Review*, v95, n1, 437-49.
- [3] Charles, Kerwin Kofi and Erik Hurst (2003). “The Correlation of Wealth Across Generations”. *Journal of Political Economy*, v111, n6, 1155-1183.
- [4] Checchi, Daniele e Luca Flabbi (2007). “Intergenerational Mobility and Schooling Decisions in Germany and Italy: The Impact of Secondary School Tracks.” *IZA Discussion Paper* No. 2876.
- [5] Cogneau D.; MAURIN E. “Parental Income and School Attendance in a Low-Income Country: A Semi-Parametric Analysis.” Document de travail. Paris : CREST, No. 2001-08, 2001.
- [6] Cunha, Flavio e James Heckman (2007). “The Technology of Skill Formation”. *NBER Working Paper* 12840
- [7] Emerson, Patrick M. e André Portela Souza (2007). “Is Child Labor Harmful? The Impact of Working Earlier in Life on Adult Earnings”. *Texto para discussão*.
- [8] Ferreira, Sergio Guimarães e Fernando A. Veloso (2003). “Mobilidade Intergeracional de Educação no Brasil”. *Pesquisa e Planejamento Econômico*, v33, n3, 481-513.
- [9] Heineck, Guido e Regina T. Riphahn (2007). “Intergenerational Transmission of Educational Attainment in Germany: The Last Five Decades.” *IZA Discussion Paper* No. 2985.
- [10] Machado, Danielle Carusi e Gustavo Gonzaga (2007). “O impacto dos fatores familiares sobre a defasagem idade-série de crianças no Brasil”. *Revista Brasileira de Economia*, v.61, p.449-476.
- [11] Maurin, E (2002). “The Impact of Parental Income on Early Schooling Transitions, a Re-examination Using Data over Three Generations”. *Journal of Public Economics*, No. 85, 2002.
- [12] Oreopoulos, Philip, Marianne Page e Ann Stevens (2006). “Does Human Capital Transfer from Parent to Child? The Intergenerational Effects of Compulsory Schooling”. *Journal of Labor Economics*, 24(4), 729-760.
- [13] Souza, André Portela e Andrew W. Horowitz (2004). “The Dispersion of Intra-Household Human Capital Across Children: A Measurement Strategy and Evidence”. *Econometric Society 2004 Latin American Meetings 176, Econometric Society*.
- [14] Souza, André Portela (2007). “Does The Gradient Matter? Further Understanding the Intergenerational Transmission of Human Capital.” *Texto para discussão*.