6

Resultados das Estimações com Variáveis Instrumentais

6.1.

Replicando os Resultados Encontrados na Literatura

Como um primeiro exercício, vamos reestimar a relação que Machado e Gonzaga descrevem, com algumas modificações já em direção àquele modelo que temos em mente. Fundamentalmente, não vamos utilizar a diferença entre anos de estudo de pais e avós como instrumento, uma vez que em um modelo de transmissão intergeracional imperfeita de atributos familiares (Becker e Tomes, 1986) essa diferença de escolaridade poderá estar correlacionada com fatores presentes no erro da equação de educação do filho. Essa correlação cresce com a existência de imperfeição no mercado de capitais enfrentado pelos avós e pais, já que os níveis de escolaridade adquiridos poderiam estar sendo determinados por essas restrições. Portanto, dada a perspectiva teórica adotada no trabalho, esse instrumento poderia não satisfazer a condição de exogeneidade necessária.

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.

Desse modo, o modelo por trás das estimações desta seção é bastante similar àquele de Machado e Gonzaga, com atributos familiares não-observados afetando concomitantemente educação de avós, pais e filhos e renda dos pais; esses atributos, no entanto, não constituirão efeitos fixos familiares, mas serão transmitidos a uma taxa λ entre 0 e 1.

A instrumentalização da educação deve retirar os efeitos indiretos desses 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. Da mesma forma, ao instrumentalizar a renda familiar per capita, estamos

expurgando vieses desses fatores e encontrando o real efeito de uma mudança no nível de renda, *ceteris paribus*, sobre o desempenho do filho. Nesse caso, o sentido do viés é incerto. Em países desenvolvidos, a renda familiar talvez não tenha um efeito causal significativo na defasagem idade-série do filho. Entretanto, no contexto de um país subdesenvolvido, pode-se esperar que o aumento de renda tenha um efeito causal positivo e significativo sobre o desempenho do aluno, se estiver relacionado à diminuição de restrições ao investimento em capital humano dos filhos e ao aumento da renda permanente da família.

Os resultados de 1º estágio da projeção de anos de estudo da mãe, do pai e renda familiar *per capita* nos instrumentos estão nas colunas da Tabela 8. Como vemos, os F-parciais confirmam a relevância das variáveis escolhidas.

<i>var. dependente:</i>	<i>anos de estudo da mãe</i>	<i>anos de estudo do pai</i>	<i>renda familiar per capita</i>
esc7_mae	-0.244 (1.62)	-0.026 (0.19)	0.117 (2.32)**
esc11_mae	0.109 (0.54)	0.034 (0.19)	-0.034 (0.50)
esc15_mae	0.000 (0.00)	-0.211 (1.73)*	-0.187 (4.17)***
prof7_mae	-0.205 (3.68)***	-0.076 (1.81)*	-0.033 (2.36)**
prof11_mae	-0.064 (1.22)	-0.146 (2.98)***	-0.013 (0.94)
prof15_mae	0.234 (6.68)***	0.209 (6.12)***	0.050 (4.80)***
esc7_pai	0.254 (1.64)*	0.325 (2.30)**	0.185 (3.69)***
esc11_pai	0.199 (0.96)	0.227 (1.18)	-0.046 (0.68)
esc15_pai	0.050 (0.35)	-0.101 (0.75)	-0.104 (2.22)**
prof7_pai	0.003 (0.06)	-0.011 (0.23)	0.015 (0.99)
prof11_pai	-0.021 (0.34)	-0.043 (0.73)	-0.038 (2.24)**
prof15_pai	0.125 (3.20)***	0.220 (5.69)***	0.047 (4.25)***
nasc_pai_1933 a 1956	0.122 (2.06)**	0.109 (1.88)*	0.055 (2.85)***
nasc_mae_1933 a 1956	-0.302 (4.35)***	-0.264 (4.18)***	0.109 (4.93)***
Obs	42439	42439	40332
R-2	0.36	0.41	0.33
F-Parcial	26.02	55.24	55.49
Prob>F	0.00	0.00	0.00

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, 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

Comparativamente aos valores dos F-parciais reportados em Machado e Gonzaga (cerca 4.000 para anos de estudo de mãe e pai e 400 para a renda per capita familiar), os do presente trabalho são bem menores. 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. No entanto, veremos abaixo que o 2º estágio pouco se altera.

O resultado de 2º estágio está na Tabela 9, ao lado de uma comparação com os resultados de Gonzaga e Machado. Qualitativamente, os resultados são idênticos: encontramos que a educação da mãe afeta positivamente o desempenho dos filhos, a educação do pai não tem efeito significativo e a renda familiar per capita se mostra altamente significativa.

Quantitativamente, os resultados também são semelhantes: enquanto Gonzaga e Machado concluem que um ano a mais de estudo da mãe reduz a probabilidade de o filho ter defasagem idade-série positiva em 1,1%, nosso resultado sugere um efeito maior, de 3,2%, porém estimado de forma mais imprecisa: nosso erro-padrão é cinco vezes maior. Como resultado, o p-valor da hipótese nula de ausência de efeito causal de educação da mãe aumenta, mas continua abaixo de 5%. A estimativa pontual sob esse método é menor que as estimativas por MQO²², 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 estimam um efeito entre 2% e 4%.

O efeito causal da renda familiar per capita é virtualmente igual nos dois estudos, e o coeficiente é maior que nas estimativas por MQO. A explicação apontada por Machado e Gonzaga é intuitiva²³, mas não parece que o aumento do efeito da renda esteja influenciado necessariamente pelo poder explicativo da variação de escolaridade entre pais e avós. Como não estamos controlando por dummies anuais, os instrumentos que usamos também podem estar correlacionados com aumentos na renda permanente da maioria das famílias

²² Também estimamos as especificações do capítulo 5 com variável dependente binária. Os resultados são qualitativamente idênticos, à exceção de que não conseguimos capturar relações não-lineares significantes. Os valores das estimativas pontuais são maiores que sob o método 2SLS.

²³ No capítulo 2.

brasileiras durante o século XX, aumentos que certamente estariam correlacionados com o aumento da renda familiar per capita medida pela pesquisa, e que devem gerar fortes efeitos sobre o desempenho escolar das crianças. Alternativamente, nossos instrumentos podem estar captando aumentos na oferta de escolas correlacionados com aumentos na renda per capita dos estados e também na renda familiar per capita. Por último, a renda familiar per capita informada na pesquisa pode ser uma medida particularmente imprecisa, e por isso sua instrumentalização também nos daria uma estimativa pontual maior.

Portanto, os resultados encontrados pelos autores parecem ser robustos à retirada daquele instrumento construído pela diferença de educação de pais e avós.

Efeitos estimados	Firmo (2008)	Machado e Gonzaga (2007)
anos de estudo da mãe	-0.032 (0.016)**	-0.011 (0.003)***
anos de estudo do pai	-0.005 (0.013)	-0.003 (0.004)
log da renda familiar per capita	-0.17 (0.020)***	-0.168 (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 9

Nas secções seguintes buscamos analisar especificamente o papel causal da escolaridade dos pais na mobilidade intergeracional de capital humano no Brasil. Também procuramos desenvolver análises que produzam estimativas comparáveis às dos estudos feitos para os países desenvolvidos, com o objetivo de avaliar se as conclusões tiradas naqueles ambientes podem ser generalizadas para o contexto de um país menos desenvolvido e com baixa escolaridade média.

6.2. 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 5 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.

6.2.1. Efeito da Mãe

Assim como no capítulo anterior, iremos estimar efeitos de pai e mãe separadamente, a princípio. A Tabela 10 mostra a projeção da educação da mãe nas variáveis escolhidas como seus instrumentos, para a amostra com todos os filhos.

var dependente: anos de estudos da mãe

	<i>PNAD96</i>	<i>PNAD88</i>
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
F-Parcial	68.36	62.57
Prob>F	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 10

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 11 mostra, para os dados da PNAD96, a matriz de correlações das variáveis utilizadas como instrumentos²⁴. 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 escola²⁵. 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.

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 11

A Tabela 12 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.

²⁴ As correlações para os dados de 1988 são quase idênticas.

²⁵ Emerson e Souza (2006) dão exemplos de que, aparentemente, estados realmente adotam políticas distintas quanto a essas duas dimensões.

especificação básica

	FILHOS		FILHO		FILHA	
	96	88	96	88	96	88
PNAD						
anos de estudo da mãe	-1,224	-1,735	-1,574	-1,848	-0,879	-1,575
	(0,449)***	(0,499)***	(0,590)***	(0,612)***	(0,582)	(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 12

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.

Em relação aos coeficientes estimados no capítulo anterior, os coeficientes de IVs são menores²⁶ que os de MQO para os dados de 1996 e 1988²⁷, como prevê a teoria²⁸. 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.

²⁶ Por coeficiente menor, entenda-se “menor em valor absoluto”, sempre.

²⁷ As idades dos pais influenciam muito pouco os coeficientes estimados através do método de MQO. Embora as especificações estimadas no capítulo anterior tenham essas variáveis entre os controles, todas as comparações entre os coeficientes estimados através de 2SLS-IV e MQO são robustas à retirada dessas variáveis.

²⁸ Embora o sentido do viés de fatores não-observados familiares omitidos sob o método MQO seja positivo, é possível que as estimativas pontuais através do método de 2SLS-IV sejam maiores que as MQO. Para isso basta que o viés de atenuação causado pela mensuração incorreta dos níveis de educação paterno e materno seja maior que aquele causado pelos fatores omitidos. Ver Oreopoulos et al. (2006).

6.2.2. Efeito do Pai

Faremos o exercício análogo para o efeito da instrução do pai. A Tabela 13 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

	<i>PNAD96</i>	<i>PNAD88</i>
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
F-Parcial	69.39	53.56
Prob>F	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 13

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).

especificação básica

	FILHOS		FILHO		FILHA	
	96	88	96	88	96	88
anos de estudo do pai	-1,287	-1,349	-0,839	-1,968	-1,809	-0,693
	(0,470)***	(0,541)**	(0,589)	(0,660)***	(0,627)***	(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

Os resultados diferentes entre os anos podem estar refletindo mudanças em algumas estruturas, mas devemos pesquisar mais para maior compreensão²⁹. Além

²⁹ Dentre as possíveis causas de tal descasamento de resultados, podemos citar também (i) a criação de um novo estado, o Tocantins (presente em 1996, mas não em 1988), que irá influenciar nos instrumentos; (ii) diferenças entre as pesquisas (ainda não muito claras); (iii) alterações nas

disso, como demonstrado no capítulo 3, a inclusão de somente um dos pais muito provavelmente está viesando nossos resultados. A próxima seção trata de tentar separar esses efeitos.

6.2.3. 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.

estruturas familiares e domiciliares que, junto com matching no mercado de casamentos podem influenciar os resultados; (iv) e o crescimento da participação feminina no mercado de trabalho.

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
F-Parcial	40.00	29.89	37.88	28.23
Prob>F	0.00	0.00	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 15

No entanto, a análise de cada instrumento individualmente mostra resultados não esperados. 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), como citado no capítulo 2, 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=42334)

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	-0,654 (1,473)	0,324 (2,071)	-1,562 (1,632)	1,083 (2,022)	0,013 (1,735)	-2,354 (2,528)
anos de estudo do pai	-0,65 (1,547)	-1,96 (2,199)	0,405 (1,693)	-3,212 (2,201)	-1,45 (1,808)	1,207 (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 instrumentos³⁰ 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.

6.2.4. 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,

³⁰ A alta correlação entre os instrumentos que indicam o nascimento antes da reforma educacional de 1971 também é indicada na correlação entre as idades dos pais: $p=0,69$.

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	-1,598	-1,256	-1,595	-0,950	-1,594
	(0,042)***	(0,051)***	(0,054)***	(0,065)***	(0,055)**	(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

<i>PNAD</i>	<i>96</i>	<i>88</i>
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
F-Parcial	56,6	43,65
Prob>F	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	-0,652 (0,212)***	-0,776 (0,249)***	-0,605 (0,268)**	-0,967 (0,309)***	-0,700 (0,286)**	-0,622 (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).

6.3. 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								
<i>var. dependente:</i>	<i>anos de estudo da mãe</i>				<i>anos de estudo do pai</i>			
<i>especificações:</i>	<i>basica</i>	<i>idades</i>	<i>idades^2</i>	<i>completa</i>	<i>basica</i>	<i>idades</i>	<i>idades^2</i>	<i>completa</i>
<i>esc7_mae</i>	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)
<i>esc11_mae</i>	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)
<i>esc15_mae</i>	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)
<i>prof7_mae</i>	-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)
<i>prof11_mae</i>	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)
<i>prof15_mae</i>	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)
<i>esc7_pai</i>	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)
<i>esc11_pai</i>	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)
<i>esc15_pai</i>	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)**
<i>prof7_pai</i>	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)
<i>prof11_pai</i>	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)
<i>prof15_pai</i>	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)
<i>nasc_pai_1933_1956</i>	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)
<i>nasc_mae_1933_1956</i>	-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
F-Parcial	40.00	3.00	0.65	0.33	37.88	1.81	2.24	1.63
Prob>F	0.00	0.00	0.82	0.99	0.00	0.03	0.00	0.06

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

A Tabela 22 mostra o mesmo padrão dos F-parciais dos instrumentos para os dados de 1988.

PNAD88								
var. dependente:	anos de estudo da mãe				anos de estudo do pai			
especificações:	basica	idades	idades^2	completa	basica	idades	idades^2	completa
esc7_mae	0.390 (2.19)**	0.150 (0.76)	0.211 (1.06)	0.253 (1.19)	0.032 (0.19)	-0.130 (0.68)	-0.102 (0.53)	-0.169 (0.81)
esc11_mae	-0.157 (0.72)	-0.324 (1.45)	-0.312 (1.39)	-0.239 (1.01)	-0.102 (0.49)	-0.219 (1.04)	-0.209 (0.99)	-0.247 (1.08)
esc15_mae	0.408 (2.25)**	0.207 (1.09)	0.186 (0.98)	0.052 (0.26)	0.427 (2.62)***	0.276 (1.62)	0.268 (1.56)	0.320 (1.74)*
prof7_mae	-0.073 (1.31)	-0.097 (1.68)*	-0.094 (1.63)	0.011 (0.15)	-0.051 (0.83)	-0.064 (0.96)	-0.063 (0.95)	-0.032 (0.40)
prof11_mae	0.093 (1.81)*	0.084 (1.64)	0.097 (1.89)*	0.057 (1.01)	0.034 (0.65)	0.030 (0.58)	0.036 (0.69)	0.082 (1.40)
prof15_mae	0.094 (2.66)***	-0.007 (0.16)	0.009 (0.22)	-0.028 (0.57)	0.106 (2.81)***	0.027 (0.63)	0.035 (0.80)	-0.010 (0.19)
esc7_pai	0.168 (0.89)	-0.273 (1.29)	-0.293 (1.38)	-0.234 (1.01)	0.378 (2.14)**	-0.025 (0.13)	-0.024 (0.12)	0.108 (0.50)
esc11_pai	0.262 (1.16)	0.106 (0.46)	0.069 (0.30)	-0.063 (0.24)	0.068 (0.32)	-0.085 (0.39)	-0.088 (0.40)	-0.256 (1.02)
esc15_pai	0.040 (0.22)	-0.155 (0.77)	-0.141 (0.70)	-0.193 (0.88)	0.329 (1.88)*	0.127 (0.66)	0.133 (0.69)	0.151 (0.70)
prof7_pai	0.039 (0.58)	-0.058 (0.85)	-0.059 (0.87)	0.030 (0.35)	-0.018 (0.31)	-0.106 (1.74)*	-0.111 (1.79)*	-0.034 (0.43)
prof11_pai	0.095 (1.38)	0.050 (0.70)	0.032 (0.44)	-0.063 (0.78)	0.115 (1.78)*	0.072 (1.09)	0.072 (1.05)	-0.025 (0.31)
prof15_pai	0.049 (1.02)	-0.013 (0.25)	-0.027 (0.50)	-0.042 (0.66)	0.061 (1.34)	-0.001 (0.01)	-0.000 (0.01)	0.006 (0.09)
nasc_pai_1933_1956	0.224 (3.41)***	0.042 (0.58)	0.118 (1.13)	0.200 (0.18)	0.164 (2.45)**	0.001 (0.01)	-0.008 (0.08)	0.579 (0.54)
nasc_mae_1933_1956	-0.042 (0.63)	-0.078 (1.14)	-0.188 (1.84)*	-0.261 (0.26)	0.049 (0.75)	0.025 (0.37)	-0.040 (0.42)	0.463 (1.50)
Obs	28308	28308	28308	28308	28308	28308	28308	28308
R-2	0.37	0.38	0.38	0.38	0.44	0.44	0.44	0.44
F-Parcial	29.89	1.17	1.45	0.61	28.23	0.78	0.81	0.99
Prob>F	0.00	0.29	0.12	0.86	0.00	0.69	0.66	0.46

fonte: 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, 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 22

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 possível³¹.

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.

³¹ Para muitos anos, mas não todos, há dados desagregados entre a capital e o interior dos estados. Para vários anos, também, há separadamente números de escolas públicas e privadas.