

4 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.

4.1. 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 mãe¹⁴.

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

¹⁴ A condição ‘pai’ ou ‘mãe’ é informada pelo entrevistado. Possivelmente não se trata de pai ou mãe biológico de uma ou mais crianças, mas essa é uma restrição à análise que não temos como superar.

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

As Tabelas A.1 e A.2 do Apêndice mostram as estatísticas descritivas das principais variáveis utilizadas, para as PNADs 1996 e 1988, respectivamente¹⁵. Destacam-se o aumento das médias de anos de estudo completos por mãe e pai, de cerca de 15% entre 1988 e 1996. Ainda assim, os valores são bastante baixos, refletindo o lento progresso educacional do Brasil durante o século passado. Além disso, a defasagem idade-série cai muito pouco no período. Vemos também que todos os valores assumidos pelas variáveis *proxies* para oferta educacional (número de escolas e de professores) aumentam.

A Figura 1 do Apêndice mostra a evolução da defasagem idade-série média de meninos e meninas, para os dados da PNAD96. São claras duas características: meninos têm defasagem idade-série média maior que meninas e essa diferença aumenta até cerca de 14 anos de idade, quando se estabiliza.

¹⁵ Não mostramos nas tabelas as dummies de coorte, de UF de nascimento de pais e filhos e de escolaridade dos avós. A razão é que as estatísticas descritivas dessas cerca de 200 dummies são muito pouco informativas.

4.2. 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.

4.2.1. 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, respectivamente¹⁶.

Esse grupo de instrumentos já foi utilizado por Emerson e Souza (2007) e Machado e Gonzaga (2007)¹⁷. 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 educação¹⁸.

¹⁶ A distribuição educacional do Brasil mostra claramente a importância prática desses ciclos escolares: há 'picos' de freqüência aos 4, 8 e 11 anos de estudo, correspondentes aos fins de cada ciclo.

¹⁷ Apesar de públicos, os dados disponíveis no site do IBGE estão consideravelmente desorganizados. Gostaríamos de agradecer enfaticamente a André Portela de Souza, que gentilmente nos cedeu as séries organizadas.

¹⁸ Machado e Gonzaga (2007) levantam a hipótese de endogeneidade desses instrumentos, no sentido de que (i) se eles estão correlacionados com a oferta educacional enfrentada pelos pais em certa UF, (ii) se os pais não migraram e (iii) se há uma correlação temporal nessa oferta educacional, é provável que esses instrumentos estejam correlacionados com a oferta educacional

4.2.2. 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 contrário¹⁹.

que seus filhos irão enfrentar, e por isso afetem diretamente o desempenho escolar da criança, violando a hipótese de exogeneidade. Os autores prosseguem a análise considerando a hipótese remota, dadas as grandes mudanças no sistema escolar desde então.

¹⁹ Ver Machado e Gonzaga (2007), que utilizam a mesma estratégia.