

FILHOS E RENDA FAMILIAR: UMA APLICAÇÃO DO EFEITO QUANTÍLICO DE TRATAMENTO

Ana Katarina Campêlo
Do Departamento de Economia da UFPE
Everton Nunes da Silva
Da UFRGS

Pesquisas sobre as conseqüências da decisão de ter filhos na renda e na oferta de trabalho são complicadas pelo fato de a fecundidade ser uma variável endógena, o que pode introduzir um viés no estimador convencional. Para estimar o efeito daquela decisão de forma mais adequada, usamos neste artigo a abordagem recentemente proposta por Abadie, Angrist e Imbens (1998 e 2002), denominada efeito quantílico de tratamento (EQT). Esse estimador utiliza a preferência dos pais por ter filhos de sexos diferentes como uma variável instrumental, permitindo variações nas estimativas ao longo dos quantis da distribuição da renda familiar. O EQT pode ser usado para determinar como uma intervenção afeta a distribuição da variável resposta para indivíduos cuja característica de reação ao tratamento é afetada pela variável instrumental. Essa técnica também tem a vantagem de ser consistente com os resultados obtidos pela regressão quantílica convencional quando o tratamento é exógeno. As estimativas do EQT com os dados da PNAD para o Brasil revelam que há diferenças importantes, entre os diversos quantis, no efeito da renda familiar sobre a decisão de ter filhos. Os resultados apontam para uma redução que varia entre 14% e 18% nos quantis considerados da distribuição de renda, sendo maior naqueles que correspondem aos extremos da distribuição de renda, em especial nos que correspondem às famílias de menor renda.

1 INTRODUÇÃO

A decisão de ter filhos é uma questão que tem despertado o interesse por parte dos pesquisadores desde a segunda metade do século passado, pelas implicações socioeconômicas e demográficas que ocasiona. Com os trabalhos de Leibenstein (1957) e Becker (1960), originou-se uma nova corrente na teoria econômica que mais tarde foi denominada Teoria da Família.¹ Essa corrente defende que pertence à família a essencial decisão de gerar e criar filhos. Seguindo o raciocínio dessa abordagem, uma vez que se opta por ter filhos, há, implicitamente, um comprometimento moral por parte dos pais de dedicarem uma determinada quantia de dinheiro e tempo para a geração e criação dos mesmos, o que pode ocasionar mudanças em seus hábitos e alterações na oferta de trabalho. Essas modificações podem provocar diferenciais nos rendimentos entre os casais que possuem e os que não possuem filhos. O tamanho da família pode ainda ser um fator agravante desses diferenciais. Dessa forma, tornam-se importantes estudos que visem a uma melhor compreensão da relação entre criação de filhos e oferta de trabalho dos pais, seja por motivos teóricos, práti-

1. Alguns importantes artigos dentro da Teoria da Família são: Becker e Lewis (1973), Ben-Porath (1973), Heckman (1980), De Tray (1973), Gronau (1973, 1977 e 1988), Goldin (1995), Willis (1973), Schultz (1973), entre outros.

cos ou para fins de políticas públicas. Alguns estudos revelam que as mães podem encontrar dificuldades em conciliar a criação dos filhos com a carreira profissional [ver, por exemplo, Goldin (1995)]. É possível então que a relação entre oferta de trabalho e fecundidade explique, em parte, o aumento da participação das mulheres na força de trabalho no último século, se podemos relacionar este último com a queda na taxa de natalidade [Coleman e Pencavel (1993)]. Outros trabalhos apontam ainda para uma relação entre salários mais baixos para as mulheres como consequência das saídas do mercado de trabalho para criar filhos [Gronau (1988) e Korenman e Neumark (1992)].

Segundo Willis (1987), a principal contribuição de Becker (1960) refere-se à sua hipótese de que existe uma parcela dos gastos com os filhos que é endógena, devido ao fato de a utilidade dos pais ser função tanto do número de filhos, quanto do capital humano investido (CHI) nos mesmos.² Assim, cabe aos pais decidir entre aumentar o tamanho da família ou intensificar o investimento em capital humano aos filhos existentes. Segundo os dados do Instituto Brasileiro de Geografia e Estatística (IBGE), há uma nítida tendência de queda da taxa de natalidade, o que sinaliza uma inclinação a se ter famílias menores. De 1980 a 2001, essa taxa declinou de 32 para 20 nascimentos por mil habitantes. As projeções indicam uma trajetória de constante decréscimo, chegando a 14 nascimentos por mil habitantes em 2050. Uma possível explicação para esse fenômeno pode estar relacionada ao custo de oportunidade de ter filhos, que é crescente, principalmente nas grandes cidades. O aumento do custo relativo da geração e criação de filhos reflete a ascendente valorização do tempo dos pais [ver Willis (1987)]. Para o caso brasileiro, Veloso (2000) analisou como a composição da renda dos pais afeta a taxa de fecundidade e o investimento nos filhos e concluiu que o custo de oportunidade do tempo é cada vez mais levado em consideração pelos pais, o que, em muitos casos, pode representar a opção dos pais em intensificar o investimento por filho, em vez de aumentar o número de filhos. Diante desse cenário, quantificar o impacto do tamanho da família sobre a renda familiar torna-se uma questão essencial tanto para os indivíduos (os pais), na medida em que os auxilia a tomar suas decisões acerca do tamanho ótimo da família, quanto para o governo, pois essa informação pode contribuir para a construção de políticas públicas mais consistentes com a realidade familiar. Pode-se ainda mencionar a relevância das questões relativas à fecundidade para o crescimento econômico de um país.

O objetivo deste artigo é buscar um maior entendimento da relação entre o tamanho da família e a renda familiar. Em particular, será estudado o efeito

2. Becker (1960) provou que a relação entre o número de filhos (*quantity*) e o capital investido nos mesmos (*quality*) é negativa, sem fazer nenhuma suposição restritiva sobre a função de utilidade ou a função de produção da família. Sua argumentação baseia-se nos custos marginais em relação ao número de filhos e ao CHI nos mesmos. Um aumento no CHI é mais caro se existem mais filhos, pois esse aumento deverá ser estendido aos demais filhos. A mesma lógica se aplica para o aumento no número de filhos. Assim, o custo marginal não é constante.

da criação de um terceiro ou mais filhos na renda dos pais. Um terceiro filho também pode ser visto como um número limite entre famílias maiores e menores. Para tal, usar-se-á a base de dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), coletada pelo IBGE, para o ano de 1999. A contribuição deste artigo está na relevância de seu objeto de estudo, ainda pouco explorado para o caso brasileiro, além da metodologia utilizada, que engloba técnicas econométricas robustas e atuais, as quais possibilitam a realização de uma análise mais precisa e generalizada da relação de interesse. Em particular, a técnica utilizada baseia-se nos artigos de Abadie, Angrist e Imbens (1998 e 2002), os quais introduzem o efeito quantílico de um tratamento (EQT). Esse método será usado para estimar o impacto da criação de filhos na distribuição da renda familiar, aplicação esta que foi abordada anteriormente na literatura nos artigos de Angrist e Evans (1998) e Abadie, Angrist e Imbens (1998), os quais utilizaram dados para os Estados Unidos. A estimação se baseia em um novo método de variáveis instrumentais que mede esse efeito nos quantis, além de lidar também com o problema da endogeneidade, presente na relação entre renda e fecundidade. Os quantis condicionais fornecem estatísticas descritivas para as distribuições e têm sido bastante usados em estudos que visam quantificar diferenciais de renda.³ Os resultados deste estudo, de forma geral, apontam para um impacto negativo da criação dos filhos na renda dos pais. No entanto, há diferenças interessantes e significativas desse efeito ao longo dos quantis estudados, o mesmo variando dentro do intervalo 14% a 18%. O efeito é marginalmente maior nas caudas da distribuição condicional de renda, em especial na esquerda (quantis inferiores). O estudo de tais diferenças possibilita uma compreensão mais detalhada e completa da relação entre fecundidade e renda, fornecendo informações mais precisas para a tomada de decisão dos agentes econômicos e do governo.

Além desta introdução, o trabalho engloba mais quatro seções. A Seção 2 aborda a metodologia do efeito quantílico de um tratamento [desenvolvida nos artigos de Abadie, Angrist e Imbens (1998 e 2002)]. A Seção 3 apresenta os dados e estatísticas descritivas. A Seção 4 fornece e discute de forma comparativa os resultados das estimativas das diversas regressões (que englobam as técnicas de mínimos quadrados ordinários (MQO), regressão quantílica, mínimos quadrados de dois estágios e o efeito quantílico de um tratamento). A Seção 5 é reservada para as conclusões.

3. Ver, por exemplo, os artigos de Buchinsky (1994), Chamberlain (1991), e Poterba e Rueben (1994).

2 METODOLOGIA

2.1 Aspectos conceituais da abordagem

A importância de se estudar os efeitos *distribucionais* de uma intervenção é largamente aceita e difundida, mas a grande maioria das análises tem se detido sobre o efeito médio, provavelmente pela sua maior simplicidade de cálculo. Por sua vez, muitos modelos econométricos restringem o efeito de um tratamento na forma de um simples efeito de locação, onde o efeito médio captaria o impacto do tratamento em todos os quantis. No entanto, em muitas áreas de pesquisa é interessante obter o efeito de um evento em vários pontos da distribuição da variável resposta, dada a presença comum de heteroscedasticidade e assimetrias, que resulta em diferenças no efeito do tratamento ao longo dos quantis. Esse é o caso da presente aplicação, pelo caráter assimétrico da distribuição dos rendimentos e heteroscedasticidade presente nos dados. Essas assimetrias são ainda mais significativas no caso brasileiro devido à proeminente desigualdade de renda existente no país. Nesse caso, é fundamental o uso de técnicas econométricas que permitam o estudo do impacto da criação de filhos em diversos pontos da distribuição condicional dos rendimentos, dado que é esperado que esse efeito seja distinto para os diversos níveis de renda. Assim, pode-se ter uma visão mais completa da relação entre o número de filhos e a renda familiar. A técnica que possibilita tal análise é a regressão quantílica, proposta por Koenker e Bassett (1978) e uma variante da mesma será usada neste artigo, denominada “EQT” e desenvolvida por Abadie, Angrist e Imbens (1998 e 2002). Assim, será obtida uma estimativa da intervenção (no caso, o impacto de um terceiro ou mais filhos) para os quantis selecionados da distribuição condicional da renda familiar, a saber: os quartis {0,25; 0,5; 0,75} e os percentis {0,1; 0,9}.

Em realidade, o impacto distribucional de um tratamento não é difícil de ser estimado quando o experimento é aleatório e também no caso em que existe perfeita conformidade entre os resultados da alocação do experimento e a própria alocação dos grupos de tratamento e controle.⁴ A aleatoriedade do experimento permite que os resultados do grupo de tratamento possam ser comparados diretamente com os do grupo de controle. Assim, inferências causais válidas podem ser obtidas através da simples comparação das distribuições dos dois grupos. A dificuldade surge na prática quando nos deparamos com experimentos não-aleatórios ou aleatórios, mas com complacência parcial (*partial compliance*) [ver Abadie, Angrist e Imbens (1998 e 2002)], e tais problemas nem sempre têm sido tratados com a devida importância e de forma criteriosa na literatura. O instrumental que será usado neste artigo [Abadie, Angrist e Imbens (1998 e 2002)] se baseia em uma fonte de variação exógena do tratamento que possibilita o cálculo do

4. Nesse caso dizemos que há uma perfeita complacência com a alocação do tratamento.

impacto do mesmo nos *quantis condicionais* da distribuição da variável resposta de interesse, mesmo na presença de tais problemas (não-aleatoriedade e complacência parcial). O EQT lida com o problema das *dummies* endógenas e o mesmo possibilita calcular o efeito do tratamento para um subgrupo cujo *status* de tratamento é afetado pelo experimento induzido pelo instrumento.⁵ Outro atrativo da metodologia é poder entender como a intervenção afeta a distribuição condicional da variável resposta e não apenas seu efeito na média condicional.

2.2 O modelo

O instrumental proposto por Abadie, Angrist e Imbens (1998 e 2002) engloba em especial três procedimentos econométricos que serão detalhados a seguir: grupos de tratamento e controle, regressão quantílica e variáveis instrumentais.

2.2.1 Grupos de tratamento e controle

A metodologia é usada para quantificar o efeito de uma intervenção na variável de interesse. Essa intervenção é denominada “tratamento” e é representada por uma variável *dummy* cujo valor 1 indica as pessoas que sofreram a intervenção (tratadas), e o valor 0 corresponde às pessoas “não-tratadas”. As primeiras são classificadas na literatura como grupo de tratamento e as últimas, como grupo de controle. No presente estudo a característica que representa o tratamento é a presença de pelo menos um terceiro filho, então nossa *dummy* (endógena), representada por *D* ao longo deste artigo, toma o valor 1 para as famílias que têm pelo menos três filhos e 0 para aquelas que têm apenas dois filhos, o que vai nos possibilitar estudar o efeito na renda familiar resultante da geração e criação de um terceiro ou mais filhos.

Para isolar o efeito de um tratamento o ideal seria comparar os resultados na *presença* e na *ausência* da intervenção para uma *mesma* pessoa. Assim, teríamos apenas o efeito do tratamento, pois estaríamos mantendo as demais características constantes. No entanto, somente um evento é observável: o resultado *com* a intervenção *ou* o resultado *sem* a intervenção. Isto é, o resultado *contrafactual* é inobservável. Para ilustrar tal fato, suponha que a intervenção seja um treinamento e que se deseje avaliar o efeito do treinamento no futuro salário dos participantes. O efeito do tratamento seria a diferença entre o salário do participante do treinamento e o salário que tal pessoa teria caso não tivesse participado do treinamento (não-observado). Como a situação *contrafactual* não é observada, torna-se mais difícil a tarefa de se conseguir isolar o efeito do tratamento. Para tal, usa-se o grupo de controle. No exemplo, o salário das pessoas do grupo de controle (as pessoas que não participaram do treinamento) servirá como estimativa para o salário das pessoas no

5. Abadie, Angrist e Imbens (1998 e 2002) referem-se a esse grupo como os *compliers*.

grupo de tratamento na situação contrafactual: salário que eles teriam caso não tivessem participado do treinamento.

Para que esse procedimento de grupos de controle e tratamento seja válido, é necessário que a *alocação* do tratamento seja *aleatória*. O experimento aleatório garante que os grupos de tratamento e controle representam uma amostra aleatória da população e não grupos com características particulares. Assim, a alocação de tais grupos será independente dos resultados em potencial. Quando isso acontece, dizemos que a alocação do tratamento é *ignorável*. Caso contrário, se o experimento não for aleatório, os resultados da simples comparação entre esses grupos serão viesados. Usando o nosso exemplo do treinamento, suponha que a alocação não tenha sido aleatória e que pessoas com maior grau de escolaridade tenham sido alocadas para o grupo de tratamento. Nesse caso, se for observado um maior salário para as pessoas treinadas, parte do valor do mesmo pode ser devido ao fato de o nível educacional ser mais elevado para os indivíduos alocados para esse grupo e não somente um resultado do treinamento, então o efeito do treinamento computado seria viesado.

Na prática, por vezes, não é fácil encontrar um grupo de controle adequado, o que também gera viés. Para vermos tal questão mais claramente, denotemos por Y nossa variável resposta e D a *dummy* de tratamento, em que $D = 1$ indica pessoas tratadas e $D = 0$, pessoas não-tratadas. Agora podemos construir a variável Y_D , que indica o resultado potencial quando $D = d$. A estimativa do efeito do tratamento com o uso da técnica de grupos de tratamento e controle pode então ser escrita como:

$$E[Y_1|D=1] - E[Y_0|D=0]$$

que indica a diferença entre os valores da variável resposta para os grupos de tratamento e controle, respectivamente. Pode-se ainda reescrever essa diferença como [ver Abadie, Angrist e Imbens (1998)]:

$$\{E[Y_1|D=1] - E[Y_0|D=1]\} + \{E[Y_0|D=1] - E[Y_0|D=0]\}$$

em que o primeiro termo entre chaves indica o efeito do tratamento para os tratados (o que desejamos calcular), enquanto o segundo termo é o viés, que surge quando a *proxy* provinda do grupo de controle (última expectância) não consegue aproximar bem o resultado contrafactual do grupo de tratamento (penúltima expectância). No exemplo do treinamento, se o salário do grupo de controle não se aproximar muito do salário que as pessoas treinadas teriam caso não tivessem participado do treinamento, teremos viés.

Esse problema pode ser contornado usando a técnica de variáveis instrumentais para um grupo de indivíduos que denominaremos *compliers*, que são os indivíduos cujo *status* de tratamento é afetado pelo instrumento, como será detalhado a seguir.

2.2.2 Variáveis instrumentais

A teoria prevê que as variáveis fecundidade e renda são determinadas conjuntamente.

Essa endogeneidade pode ser percebida na literatura afim que usa tanto modelos da renda em função do número de filhos como a relação inversa, isto é, invertendo a variável dependente com a co-variável. A presença de endogeneidade nessa relação impossibilita a estimação de efeitos causais válidos e ignorar tal problema levaria à obtenção de estimativas viesadas (por exemplo, quando do uso de mínimos quadrados ou de regressão quantílica básica).

Uma possível solução para o problema da endogeneidade é utilizar variáveis instrumentais. Para tal, deve-se encontrar uma variável, denominada *instrumento* (neste artigo representada pela variável Z), que seja correlacionada com a co-variável e não-correlacionada com a variável resposta. Uma vez encontrado esse instrumento, os parâmetros do modelo podem ser estimados consistentemente, pois a nova co-variável (o instrumento) agora é exógena, isto é, determinada aleatoriamente com relação à variável resposta. A grande questão é, então, encontrar um bom instrumento, o que não é uma tarefa fácil em algumas aplicações.

Neste artigo será utilizada uma variável instrumental (VI) construída a partir da preferência dos pais por ter filhos de ambos os gêneros (masculino e feminino). O instrumento, nesse caso, refere-se à composição do sexo dos dois primeiros filhos como um indicador para a geração de um terceiro filho. Famílias que têm os dois primeiros filhos do mesmo gênero apresentam uma probabilidade maior de gerar um terceiro filho que aquelas que já tiveram um menino e uma menina (sendo estes percentuais 43,64% e 38,29%, respectivamente). Para os Estados Unidos (dados de 1980) esses números são, respectivamente, 43,2% e 37,2% [ver Angrist e Evans (1998)].

Na literatura relacionada temos o artigo de Ben-Porath e Welch (1976), que contribuiu para evidenciar que o crescimento populacional pode ser explicado, em parte, pela preferência dos pais no que concerne ao sexo dos filhos. Segundo os autores, a preferência dos pais por ter filhos de ambos os sexos pode se dar por dois motivos: *a*) pelo próprio desejo em ter um menino e uma menina; e *b*) pelo conflito de opiniões, isto é, os pais divergem em relação à preferência pelo sexo dos filhos, o que ocasiona ao casal, como unidade familiar, o desejo de ter filhos de ambos os gêneros. Essa hipótese foi comprovada

empiricamente pelos autores, que concluem que o casal que não obteve o resultado esperado (ambos os gêneros) é significativamente mais propenso a continuar a ter filhos.

Outra evidência desse estudo, que no mínimo parece curiosa, principalmente em sociedades que apregoam igualdade entre os sexos, refere-se à tendência dos pais de querer ao menos um filho do sexo masculino. Possíveis explicações para esse fato, segundo o artigo em questão, seriam que: *a*) meninos são esperados a contribuir mais do que meninas na renda familiar; e *b*) meninos são uma fonte mais segura para dar suporte na velhice dos pais.⁶

Dado ser o gênero do bebê um evento aleatório, é pouco provável que um indicador de mesmo sexo para os dois primeiros filhos esteja associado com a variável resposta (rendimento familiar), o que lhe concede certa atratividade como instrumento para estudar a relação entre o número de filhos e a renda familiar, por ser correlacionado com o primeiro e independente da última.⁷

Por fim, salienta-se que comumente não há uma perfeita complacência entre o *status* da *dummy* usada como instrumento e o *status* da *dummy* endógena, o que pode comprometer a qualidade da instrumentalização, isto é, há casos em que famílias com os dois primeiros filhos de mesmo gênero não têm um terceiro filho e outros em que a família já tem uma menina e um menino e um terceiro filho é gerado. Nesses casos dizemos que o *status* do instrumento (Z) não induz o *status* da *dummy* endógena (D), ou seja, $Z \neq D$. Imbens e Angrist (1994) denominaram *compliers* os demais casos, para os quais temos $Z = D$. Em resumo, os *compliers* representam as famílias que gerariam um terceiro filho se os dois primeiros fossem do mesmo gênero (se $Z = 1$, então $D = 1$) e não teriam um terceiro filho se os dois primeiros fossem de ambos os sexos (se $Z = 0$ então $D = 0$). Note-se, porém, que não é possível identificar os *compliers*, pois, novamente, observamos apenas uma das seguintes situações: *ou* a família tem os dois primeiros filhos do mesmo sexo *ou* os tem de gêneros opostos. O artigo de Imbens e Angrist (1994), no entanto, propõe um procedimento para se isolar o efeito do tratamento sobre a subpopulação dos *compliers*, que será descrito na próxima seção. Esse procedimento não visa extrair o subgrupo dos *compliers* da população, mas sim calcular o efeito do tratamento para esse subgrupo usando a totalidade dos dados. A identificação do efeito do tratamento sobre os *compliers* será possível com o auxílio de pesos, cuja construção veremos adiante.⁸

6. Essas possíveis diferenças nos custos ou benefícios associados a meninos e meninas são observadas, em maior grau, em países em desenvolvimento [Ben-Porath e Welch (1976)].

7. Entre os estudos que usaram tal instrumento estão os seguintes: Abadie, Angrist e Imbens (1998) e Angrist e Evans (1998).

8. Imbens e Angrist (1994) classificaram duas outras subpopulações que auxiliarão no processo de identificação dos *compliers*: as famílias *always-takers* e as *never-takers*. As *always-takers* são as famílias que sempre têm o terceiro filho, independentemente do sexo dos dois primeiros filhos (isto é, $D = 1$ se $Z = 0$ ou $Z = 1$). E as *never-takers* nunca geram um terceiro filho (ou seja, $D = 0$ se $Z = 1$ ou $Z = 0$).

2.2.3 A identificação dos *compliers* e o efeito médio local de tratamento

O EQT para *dummies* endógenas [Abadie, Angrist e Imbens (1998 e 2002)] usado neste artigo é uma generalização do efeito médio local de um tratamento (EMLT) [Imbens e Angrist (1994)]. O último concentra-se no efeito do tratamento na média condicional, enquanto o primeiro é uma extensão para os quantis (na seção seguinte abordaremos o EQT e a técnica de regressão quantílica).

Para identificar o efeito do tratamento para os *compliers*, Imbens e Angrist (1994) acrescentam às tradicionais hipóteses da técnica de variáveis instrumentais⁹ a hipótese de monotonicidade. Esta última hipótese diz que o instrumento só pode afetar a *dummy* em apenas uma direção. Na aplicação deste artigo em particular, a presença dos dois primeiros filhos de mesmo sexo ($Z = 1$) afetaria a *dummy* endógena apenas na direção de aumentar a probabilidade de ter um terceiro filho. Defina D_z como sendo o valor da *dummy* endógena dado $Z = z$. A hipótese de monotonicidade pode então ser escrita como:

$$P(D_1 \geq D_0 | X) = 1$$

Imbens e Angrist (1994) mostram que, dadas as hipóteses descritas no parágrafo anterior, tem-se o seguinte teorema:

Teorema 1

$$\frac{E[Y|Z=1, X] - E[Y|Z=0, X]}{E[D|Z=1, X] - E[D|Z=0, X]} = E[Y_1 - Y_0 | X, D_1 > D_0]$$

O parâmetro identificado pelo Teorema 1 é denominado EMLT para os *compliers*, grupo descrito pela condição $D_1 > D_0$, que implica $Z = D$. O conceito dos *compliers* é a base fundamental no processo de obtenção do EMLT e possibilita uma explicação simples para o fato de a técnica de variáveis instrumentais funcionar nesse caso. Para esse subgrupo, vimos que $D = Z$, então, nesse caso, é possível substituir a *dummy* endógena D pelo instrumento (exógeno) Z . Podemos dizer então que, dadas as co-variáveis X , o *status* de tratamento é ignorável para os *compliers*,¹⁰ ou seja, tem-se independência entre os resultados potenciais e a alocação do tratamento para esse grupo em particular. Dessa forma, para os *compliers*, a comparação entre as médias por *status* de tratamento (diferença entre as médias condicionais dos grupos de tratamento e controle)

9. Essas hipóteses são: a) o instrumento deve ser independente dos resultados potenciais, dado a matriz de co-variáveis X ; b) a única forma de Z influenciar Y é pelo seu efeito em D ; c) a distribuição do instrumento é não-degenerada; e d) é necessário haver uma relação entre D e Z para que se possa realizar o primeiro estágio dentro da teoria de variáveis instrumentais.

10. Ver Lema 1 [Abadie, Angrist e Imbens (1998)].

nos dá o EMLT, mesmo se a alocação do tratamento não for independente dos resultados potenciais na população. Esse resultado é descrito na relação abaixo:

$$E[Y_1 - Y_0 | X, D_1 > D_0] = E[Y | D=1, D_1 > D_0, X] - E[Y | D=0, D_1 > D_0, X]$$

Para que a expectativa citada anteriormente possa ser computada na prática, é necessário isolar o subgrupo dos *compliers*, o que pode ser realizado com o auxílio da seguinte função:

$$k = k(D, Z, X) = 1 - \frac{D \cdot (1 - Z)}{1 - E[Z|X]} - \frac{Z \cdot (1 - D)}{E[Z|X]} \quad (1)$$

Note-se que k assume o valor de 1 quando $D = Z$. Para os demais casos, k assume valor negativo. Nesse caso, qualquer parâmetro definido como solução de uma condição de momento envolvendo (Y, D, X) pode ser identificado para os *compliers*, como mostra o Lema 2 (reescrito a seguir) em Abadie (1997):¹¹

Lema 2

Seja $\psi(Y, D, X)$ qualquer função real mensurável de (Y, D, X) . Então, dadas as hipóteses assumidas,¹² tem-se:

$$\frac{E[k \cdot \psi(Y, D, X)]}{P(D_1 > D_0)} = E[\psi(Y, D, X) | D_1 > D_0]$$

A partir do Lema 2, Abadie, Angrist e Imbens (1998 e 2002) desenvolvem o estimador do EQT, a ser descrito na Subseção 2.2.4.

2.2.4 Efeito quantílico de um tratamento (EQT)

O Lema 2 possibilita a generalização do EMLT para os quantis condicionais da variável resposta, como veremos a seguir. Iniciemos com a descrição da função quantil condicional da variável resposta Y dado X e D para os *compliers*:

$$Q_\theta = (Y | X, D, D_1 > D_0) = \alpha_\theta D + X' \beta_\theta \quad (2)$$

11. Conforme descrito em Abadie, Angrist e Imbens (1998), a hipótese de monotonicidade permite escrever a esperança de y dado X em termos de expectativas para os *compliers*, *always-takers* e *never-takers*, nesta ordem:

$$E[\psi | X] = E[\psi | X, D_1 > D_0] \cdot P(D_1 > D_0 | X) + E[\psi | X, D_1 = D_0 = 1] \cdot P(D_1 = D_0 = 1 | X) + E[\psi | X, D_1 = D_0 = 0] \cdot P(D_1 = D_0 = 0 | X)$$

Rearranjando termos e desenvolvendo este resultado (ver o artigo em referência para maiores detalhes) chega-se à expressão descrita no Lema 2.

12. As hipóteses em referência são as da nota de rodapé de número 9 mais a hipótese de monotonicidade.

Note-se que o parâmetro de interesse no modelo, α_θ , nos dá a diferença entre os quantis de Y_1 e Y_0 e *não* os quantis da diferença $Y_1 - Y_0$. Como salientado por Abadie, Angrist e Imbens (2002), o modelo permite saber se, por exemplo, o programa de treinamento deslocou para cima o percentil 0,10 da distribuição condicional da renda, mas não possibilita saber se as pessoas que originalmente estavam no percentil 0,10 da distribuição condicional da renda tiveram aumento em seus rendimentos.

Como mencionado, o estimador quantílico de tratamento é obtido quando a função $\psi(\cdot)$ no Lema 2 toma a forma da função perda de regressão quantílica [Koenker e Bassett (1972)], $\rho_\theta(u)$, dada por:

$$\rho_\theta(u) = u(\theta - I(u < 0)) = \begin{cases} \theta u, & \text{se } u \geq 0 \\ (\theta - 1)u, & \text{se } u < 0 \end{cases}$$

onde θ representa o quantil desejado (por exemplo, $\theta = 0,5$ representa a mediana) e u é o erro. Dessa forma, os parâmetros da função quantil em (1) podem ser obtidos minimizando-se a seguinte função perda:

$$(\alpha_\theta, \beta_\theta) = \arg \min_{(\alpha, \beta)} E[k \cdot \rho_\theta(Y - \alpha D - X' \beta)] \quad (3)$$

As estimativas de tais parâmetros são computadas usando-se a versão amostral do problema anterior:

$$(\hat{\alpha}_\theta, \hat{\beta}_\theta) = \arg \min_{(\alpha, \beta)} \sum_{i=1}^n \hat{k}_i \cdot \rho_\theta(y_i - \alpha d_i - x_i' \beta) \quad (4)$$

Abadie, Angrist e Imbens (1998 e 2002) provam a consistência e normalidade assintótica dos estimadores anteriormente definidos.¹³

Da mesma forma que os estimadores de variáveis instrumentais são equivalentes ao de mínimos quadrados quando o *status* de tratamento é exógeno, o estimador quantílico de tratamento torna-se o estimador de regressão quantílica quando não há instrumentalização. O EQT preserva então todas as vantagens do modelo de regressão quantílica:¹⁴ a) caracteriza a distribuição condicional

13. A função objetiva populacional em (3) é globalmente convexa em $(\alpha_\theta, \beta_\theta)$, mas na versão amostral em (4) o peso k é negativo quando D não é igual a Z , o que a torna não-convexa. Dessa forma, não é possível ter a representação em programação linear da função objetiva amostral, dificultando assim o processo de estimação [Abadie, Angrist e Imbens (1998)]. Para lidar com tal dificuldade, Abadie, Angrist e Imbens (2002) sugerem uma abordagem alternativa (não-paramétrica) para o cálculo de k , que é assintoticamente equivalente à quantidade em (1), mas gera apenas valores positivos para os pesos. As estimativas neste artigo seguem essa nova abordagem descrita em detalhes em Abadie, Angrist e Imbens (2002).

14. De fato, o EQT é obtido usando-se uma abordagem similar à técnica de regressão quantílica ponderada.

da variável resposta em qualquer ponto dessa distribuição e não apenas a média condicional, possibilitando assim ter-se uma idéia da forma da distribuição; *b*) a função objetivo é uma soma ponderada de desvios absolutos, fornecendo uma medida de locação robusta; *c*) pode fornecer estimadores mais eficientes que os de mínimos quadrados quando os erros não seguem a distribuição normal. O modelo de regressão quantílica pode ser interpretado de forma similar ao modelo de mínimos quadrados; no primeiro tem-se o efeito nos quantis, enquanto no segundo tem-se o efeito médio. No nosso exemplo do treinamento, a técnica de regressão quantílica permite ver o efeito do programa, por exemplo, para o quantil 0,10, pessoas de renda mais baixa e para o quantil 0,9, pessoas de renda mais alta. O retorno da participação no treinamento certamente pode ser distinto entre esses dois grupos. Assim, a regressão quantílica possibilita estudar o efeito do treinamento de forma mais completa, para as várias classes de renda, o que certamente fornece informações mais detalhadas para o processo de decisão dos agentes.

3 DADOS E ESTATÍSTICAS DESCRITIVAS

Na presente aplicação, a variável resposta (Y) é o log da renda familiar¹⁵ (excluindo-se a renda dos filhos e rendas de outras fontes), a *dummy* endógena (D) indica mulheres com três ou mais filhos (que denominamos “Mais de dois filhos” nas tabelas, na seção de descrição dos resultados) e a variável instrumental (Z) indica se os dois primeiros filhos são do mesmo sexo (ou dois meninos ou duas meninas). Entre as co-variáveis estão: *dummies* educacionais;¹⁶ *dummy* para o caso de o primeiro filho ser do sexo masculino; idade da mãe; idade da mãe quando da primeira gravidez; raça da mãe; *dummies* regionais; *dummy* metropolitana; e *dummy* área urbana. As Tabelas 1 e 2 mostram algumas estatísticas descritivas que fornecerão uma visão da composição dos dados e dos critérios de seleção da amostra.

Os dados usados neste artigo provêm da PNAD para o ano de 1999. A Tabela 1 mostra estatísticas de participação na força de trabalho de mulheres nas faixas etárias 21-35 e 36-50 e a taxa de fecundidade das mesmas. Vê-se que 53,69% das mulheres na faixa etária 21-35 estão na força de trabalho, enquanto esse percentual é de 57,12% para mulheres em idades entre 36-50 anos. A média de filhos para o grupo 21-35 é 1,45 e para o grupo 36 e 50 é praticamente o dobro (2,79).

A seleção do banco de dados baseou-se em informações sobre a oferta de trabalho, o sexo dos dois primeiros filhos das famílias e as características das

15. A renda familiar usada exclui os agregados.

16. A escolha de modelar a variável educação através de *dummies* educacionais que correspondem ao número de anos completos de estudo provém do artigo de Gonzaga e Soares (1999), o qual sugere a presença de não-linearidades no retorno à educação.

TABELA 1
ESTATÍSTICAS DE FECUNDIDADE E OFERTA DE TRABALHO

Mulheres com idade entre 21 e 35 anos

Média de filhos nascidos 1,45
Porcentagem com dois filhos ou mais 39,64
Porcentagem que trabalhou no último ano 53,69
Total de observações 44.462

Mulheres com idade entre 36-50 anos

Média de filhos nascidos 2,79
Porcentagem com dois filhos ou mais 59,57
Porcentagem que trabalhou no último ano 57,12
Total de observações 33.513

Mulheres com idade entre 21-35 anos com dois ou mais filhos

Porcentagem com mais de dois filhos 40,60
Porcentagem que trabalhou no último ano 45,73
Total de observações 17.628

Fonte: Dados da pesquisa.

mães.¹⁷ A amostra selecionada consiste em famílias nas quais as mães têm idades entre 21 e 35 anos e satisfazem os seguintes requisitos: *a*) têm pelo menos dois filhos;¹⁸ *b*) da renda familiar foi descontada a renda dos filhos e rendas de outras fontes; *c*) o primeiro filho (ou filha) tinha idade inferior a 18 anos na época da pesquisa. A amostra selecionada totalizou 13.175 famílias.

Na PNAD não é possível identificar os filhos que não habitam com os pais, então a restrição da amostra a famílias cujo primeiro(a) filho(a) tinha menos de 18 anos foi uma consequência de haver maior chance de os filhos maiores de 18 anos residirem em outro domicílio. A escolha do limite inferior da faixa etária das mães advém do fato de que poucas mulheres com idade inferior a 21 anos têm ao menos dois filhos (esse percentual é de aproximadamente 1,6% para os dados utilizados). O limite superior de 35 anos, por sua vez, foi escolhido no intuito de não se ter viés de seleção em consequência de a

17. A seleção dos dados baseou-se no artigo de Angrist e Evans (1998), tendo sido levadas em conta as considerações feitas pelos autores relativas a essa aplicação em particular. Foi também tomado como base o modelo usado por esses autores.

18. Para se construir a *dummy* de famílias que tiveram ao menos um terceiro filho, a amostra se restringiu às famílias que tinham pelo menos dois filhos, assim possibilitando estimar o efeito marginal da criação de um terceiro ou mais filhos [ver Angrist e Evans (1998)].

idade máxima do primeiro filho (ou filha) ter sido limitada a 18 anos. Para o grupo selecionado (mulheres na faixa etária 21-35 com dois ou mais filhos), o filho (ou filha) mais velho de 86,40% das mães tinha idade inferior a 18 anos. Embora essas mães possam parecer pertencentes a um grupo jovem de alta fecundidade, não usual, convém salientar que 39,64% de todas as mulheres na faixa etária 21-35 estão nesse grupo. O percentual é similar para as mulheres nessa faixa de idade com mais de dois filhos (40,60%).

A Tabela 2 consiste em estatísticas descritivas das co-variáveis, instrumento e variável resposta. A co-variável de maior interesse em nosso modelo é o indicador de famílias com mais de dois filhos e seu instrumento na aplicação será a variável “famílias com os dois primeiros filhos do mesmo sexo”. Entre as mulheres que tiveram dois filhos, 40% tiveram um terceiro filho e aproximadamente 50% tiveram os dois primeiros filhos do mesmo sexo, enquanto apenas 1% teve gêmeos na primeira gravidez.¹⁹ Vê-se ainda que 48% dos primeiros nascimentos nessas famílias foram de crianças do sexo masculino. A Tabela 2 ainda revela que a idade média das mães na amostra selecionada é de aproximadamente 30 anos e a idade média, quando da primeira gravidez, é de 20 anos. Na parte inferior da Tabela 2 tem-se as variáveis demográficas e de oferta de trabalho: a renda média das mães, a renda média da família, nível educacional e indicadores de raça, região, residência em área urbana e residência em área metropolitana. A maior parte das famílias (80%) mora em área urbana, entre as quais aproximadamente a metade mora em área metropolitana. Nota-se ainda que boa parte das famílias na amostra (36%) mora na região Nordeste, seguindo-se as regiões Sudeste (19%) e Sul (19%). É interessante também mencionar que grande parte das mães, dentro do grupo selecionado, apresenta baixa escolaridade, tendo 40% delas menos de quatro anos de estudo, enquanto apenas cerca de 1% tem pelo menos o terceiro grau completo. Em termos raciais, metade das mulheres da amostra é branca, enquanto a outra metade classifica-se como de outras etnias.

4 APLICAÇÃO

Se a fecundidade e a renda são determinadas conjuntamente, como sugere a teoria econômica [ver Browning (1992)], isso faz com que estimativas da relação entre essas variáveis pelo método de MQO ou de regressão quantílica (RQ) não tenham uma interpretação causal. A estimação por meio de variáveis instrumentais do efeito EQT contorna esse problema. Primeiramente, reportaremos resultados baseados nas técnicas tradicionais de MQO e RQ (Tabela 3) para fins de comparação e em seguida mostraremos os resultados de mínimos

19. Além do instrumento “mesmo sexo”, que indica as famílias que tiveram os dois primeiros filhos do mesmo sexo, o artigo de Angrist e Evans (1998) também utilizou o instrumento “gêmeos na primeira gravidez”. É esperado que as famílias que têm gêmeos quando da primeira gravidez da mãe têm maiores chances de ter um terceiro filho que as famílias que geraram um único filho na primeira gravidez.

TABELA 2
ESTATÍSTICAS DESCRITIVAS PARA A AMOSTRA DE MULHERES COM IDADE ENTRE 21 E 35 ANOS COM DOIS OU MAIS FILHOS, SENDO O PRIMOGÊNITO COM IDADE INFERIOR A 18 ANOS

Variáveis	Média	Desvio-padrão
Número de filhos nascidos	2,76	1,20
<i>Dummy</i> famílias com mais de dois filhos	0,40	-
<i>Dummy</i> famílias com os dois primeiros filhos do mesmo sexo	0,50	-
<i>Dummy</i> famílias com gêmeos na primeira gravidez	0,01	-
<i>Dummy</i> primeiro filho homem	0,48	-
Idade da mãe	29,57	3,93
Idade da mãe quando da primeira gravidez	20,27	3,33
Renda da mãe	137,09	380,90
Renda familiar	668,29	1.028,53
<i>Dummy</i> raça (=1 se branca)	0,48	-
<i>Dummy</i> área urbana	0,80	-
<i>Dummy</i> área metropolitana	0,37	-
<i>Dummy</i> região Nordeste	0,36	-
<i>Dummy</i> região Norte	0,10	-
<i>Dummy</i> região Sudeste	0,19	-
<i>Dummy</i> região Sul	0,19	-
<i>Dummy</i> região Centro-Oeste	0,16	-
<i>Dummy</i> < 4 anos de estudo	0,40	-
<i>Dummy</i> 4 anos de estudo	0,11	-
<i>Dummy</i> 4 < anos de estudo < 8	0,24	-
<i>Dummy</i> 8 anos de estudo	0,08	-
<i>Dummy</i> 8 < anos de estudo < 11	0,17	-
<i>Dummy</i> 11 anos de estudo	0,01	-
<i>Dummy</i> 11 < anos de estudo < 15	0,04	-
<i>Dummy</i> 15 ou mais anos de estudo	0,01	-

quadrados de dois estágios (MQ2E) e do EQT (Tabela 4), que levam em consideração o problema da endogeneidade. Os resultados das Tabelas 3 e 4 são dados X e D ; e os resultados da Tabela 4 devem ser interpretados como relacionados à subpopulação dos *compliers*.²⁰

No nosso modelo, a variável de maior interesse é a *dummy* endógena “mais de dois filhos”, a qual é instrumentalizada pela *dummy* “dois primeiros filhos do mesmo sexo” nos modelos de MQ2E e EQT.²¹ Os resultados de MQO (Tabela 3) apontam para uma redução de aproximadamente 12% no log da renda familiar como consequência da criação de filhos. A técnica de MQ2E (Tabela 4), por sua vez, indica uma queda na renda em torno de 46%, porém o coeficiente não é estatisticamente significativo ao nível de 5% (sendo apenas significativo a 10%).

As estimativas de RQ são apresentadas na Tabela 3 e revelam um efeito negativo no quantil 0,10 de 16,6%. Esse efeito negativo na renda tende a diminuir (em valor absoluto) à medida que avançamos para os quantis mais elevados, chegando a ser 10,1% no quantil 0,75. Essa tendência de queda se reverte no quantil 0,9, sendo o coeficiente nesse quantil $-0,152$. Os resultados apontam, então, para efeitos em maior magnitude tanto para as famílias mais pobres quanto as mais abastadas. É interessante mencionar que os resultados dessa aplicação, usando dados dos Estados Unidos [Abadie, Angrist e Imbens (1998)], mostram que as famílias mais pobres sofrem redução mais significativa em seus rendimentos do que as famílias mais ricas, e os resultados de RQ apontam para uma diminuição desse efeito (em valor absoluto) à medida que nos movemos para a cauda direita da distribuição. Para os Estados Unidos, no quantil 0,1 o valor do

20. Os pesos do primeiro estágio para a estimação do EQT foram calculados usando o estimador sugerido por Abadie, Angrist e Imbens (2000 e 2002):

$$\hat{k}_i(U_i) = 1 - \frac{D_i \cdot (1 - \hat{v}(U_i))}{1 - \hat{\pi}(X_i)} - \frac{(1 - D_i) \cdot \hat{v}(U_i)}{\hat{\pi}(X_i)}$$

seja $U = (Y, D, X)$ e $V = (Z, U)$, então a amostra pode ser indexada por $\{\{V_i^m\}\}_{i=1}^L$, onde $\{V_i^m\}$ são subsequências para distintos valores fixos de X . Um estimador não-paramétrico em séries polinomiais de $v = E[Z | Y, D, X]$, $\hat{v}(U)$, é dado pela projeção de MQO de $\{Z_i^m\}$ em $\{p^k(Y_i)\}$, onde $p^k(Y) = (Y^{2k(1)}, \dots, Y^{2k(K)})$, sendo $\{\lambda(k)\}_{k=1}^\infty$ uma seqüência crescente de números inteiros positivos e K um número inteiro positivo. Tem-se que $\hat{v}(U)$ são os valores ajustados resultantes e $\hat{\pi}(X)$ é um estimador obtido através da média empírica de Z para cada valor fixo de X . Na presente aplicação, uma análise exploratória mostrou que termos de iteração com D foram significativos, então v foi estimado separadamente para $D = 0, 1$ (ou seja, todos os termos na série foram iteragidos com D). Uma análise exploratória adicional mostrou que a maioria dos termos em X não foi estatisticamente significativa, então esses foram descartados. A única co-variável em X com poder explicativo, dados D e Y , foi o indicador de “primeiro filho homem”. Dessa forma, a série conteve uma constante e termos em Y , os quais foram iteragidos com D e com o indicador de “primeiro filho homem”. A seleção da ordem do polinômio foi realizada através dos critérios *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC), que são critérios de *cross-validation* baseados no valor da log-verossimilhança. Os polinômios estimados contiveram termos até a décima ordem e o melhor modelo apontado pelos critérios do AIC e do BIC foi o de ordem 1 para todos os casos resultantes das iterações.

21. Como colocado na Subseção 2.2.3, nos modelos que usam a técnica de variáveis instrumentais e para essa aplicação em particular, o instrumento (*dummy* “dois primeiros filhos do mesmo sexo”) afeta a *dummy* endógena (“mais de dois filhos”) apenas na direção de se ter um terceiro filho.

TABELA 3
RESULTADOS DE MÍNIMOS QUADRADOS ORDINÁRIOS E REGRESSÃO QUANTÍLICA

Log renda familiar	MQO	$\theta = 0,1$	$\theta = 0,25$	$\theta = 0,5$	$\theta = 0,75$	$\theta = 0,9$
Constante	4,416 (0,000)	4,056 (0,000)	4,143 (0,000)	4,405 (0,000)	4,795 (0,000)	4,994 (0,000)
Mais de 2 filhos	-0,118 (0,000)	-0,166 (0,000)	-0,154 (0,000)	-0,120 (0,000)	-0,101 (0,000)	-0,152 (0,000)
Idade	0,033 (0,000)	0,022 (0,000)	0,028 (0,000)	0,035 (0,000)	0,035 (0,000)	0,043 (0,000)
1° filho homem	0,138 (0,000)	0,299 (0,000)	0,169 (0,000)	0,102 (0,000)	0,048 (0,001)	0,053 (0,028)
Raça	0,190 (0,000)	0,114 (0,000)	0,138 (0,000)	0,169 (0,000)	0,215 (0,000)	0,214 (0,000)
Urbana	0,299 (0,000)	0,151 (0,000)	0,187 (0,000)	0,243 (0,000)	0,238 (0,000)	0,243 (0,000)
Metropolitana	0,081 (0,000)	0,073 (0,000)	0,075 (0,000)	0,084 (0,000)	0,094 (0,000)	0,088 (0,000)
Norte	0,227 (0,000)	0,188 (0,000)	0,211 (0,000)	0,215 (0,000)	0,252 (0,000)	0,295 (0,000)
Sudeste	0,409 (0,000)	0,413 (0,000)	0,461 (0,000)	0,448 (0,000)	0,402 (0,000)	0,341 (0,000)
Sul	0,313 (0,000)	0,341 (0,000)	0,379 (0,000)	0,332 (0,000)	0,252 (0,000)	0,248 (0,000)
Centro-Oeste	0,320 (0,000)	0,284 (0,000)	0,318 (0,000)	0,347 (0,000)	0,313 (0,000)	0,323 (0,000)
Idade teve 1° filho	-0,014 (0,000)	-0,015 (0,000)	-0,011 (0,000)	-0,016 (0,000)	-0,013 (0,000)	-0,015 (0,000)
Educação = 1	-0,011 (0,794)	-0,014 (0,901)	0,026 (0,511)	0,012 (0,805)	-0,086 (0,111)	-0,024 (0,795)
Educação = 2	0,009 (0,798)	0,042 (0,296)	0,034 (0,382)	0,062 (0,088)	-0,025 (0,533)	-0,005 (0,923)
Educação = 3	0,074 (0,013)	0,056 (0,111)	0,075 (0,021)	0,080 (0,019)	0,060 (0,088)	0,134 (0,001)
Educação = 4	0,144 (0,000)	0,120 (0,000)	0,120 (0,000)	0,141 (0,000)	0,121 (0,000)	0,180 (0,001)
Educação = 5	0,246 (0,000)	0,191 (0,000)	0,205 (0,000)	0,260 (0,000)	0,220 (0,000)	0,274 (0,004)
Educação = 6	0,270 (0,000)	0,225 (0,000)	0,223 (0,000)	0,265 (0,000)	0,284 (0,000)	0,357 (0,012)
Educação = 7	0,367 (0,000)	0,270 (0,000)	0,303 (0,000)	0,355 (0,000)	0,355 (0,000)	0,512 (0,000)
Educação = 8	0,528 (0,000)	0,376 (0,000)	0,446 (0,000)	0,517 (0,000)	0,513 (0,000)	0,646 (0,000)
Educação = 9	0,525 (0,000)	0,377 (0,000)	0,439 (0,000)	0,542 (0,000)	0,577 (0,000)	0,650 (0,001)
Educação = 10	0,648 (0,000)	0,456 (0,000)	0,565 (0,000)	0,606 (0,000)	0,752 (0,000)	0,818 (0,000)
Educação = 11	0,952 (0,000)	0,720 (0,000)	0,795 (0,000)	0,947 (0,000)	1,050 (0,000)	1,181 (0,000)
Educação = 12	1,230 (0,000)	1,028 (0,000)	1,053 (0,000)	1,175 (0,000)	1,342 (0,000)	1,506 (0,034)
Educação = 13	1,531 (0,000)	1,324 (0,000)	1,388 (0,000)	1,571 (0,000)	1,609 (0,000)	1,997 (0,000)
Educação = 14	1,406 (0,000)	1,032 (0,000)	1,217 (0,000)	1,436 (0,000)	1,630 (0,000)	1,683 (0,000)
Educação = 15+	1,880 (0,000)	1,639 (0,000)	1,699 (0,000)	1,877 (0,000)	1,982 (0,000)	2,063 (0,000)

Nota: Valores-p entre parênteses.

coeficiente é $-0,092$ e o mesmo vai decrescendo (em valor absoluto) ao longo dos quartis e atinge o valor de $-0,027$ no quantil 0,9. Dessa forma, os resultados brasileiros refletem a mesma tendência observada para os Estados Unidos, com a diferença de que essa tendência é revertida no caso brasileiro no quantil 0,9. Vale salientar ainda que a magnitude da queda na renda resultante de um terceiro filho é maior no caso brasileiro do que nos Estados Unidos.

Partindo para a Tabela 4, verificamos que o comportamento observado nos resultados de RQ é mantido quando usamos o EQT, isto é, o efeito nega-

TABELA 4
**RESULTADOS DE MÍNIMOS QUADRADOS DE DOIS DÍGITOS E EFEITO QUANTÍLICO DE UM
 TRATAMENTO PARA OS COMPLIERS**

Log renda familiar	MQZE	$\theta = 0.1$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$	$\theta = 0.9$
Constante	4,599 (0,000)	3,944 (0,000)	4,128 (0,000)	4,464 (0,000)	4,737 (0,000)	5,042 (0,000)
Mais de 2 filhos	-0,457 (0,067)	-0,181 (0,000)	-0,164 (0,000)	-0,143 (0,000)	-0,142 (0,000)	-0,180 (0,000)
Idade	0,044 (0,000)	0,026 (0,000)	0,031 (0,000)	0,036 (0,000)	0,039 (0,000)	0,045 (0,000)
1º filho homem	0,146 (0,000)	0,431 (0,000)	0,350 (0,000)	0,278 (0,000)	0,241 (0,000)	0,247 (0,000)
Raça	0,170 (0,000)	0,151 (0,000)	0,163 (0,000)	0,208 (0,000)	0,245 (0,000)	0,246 (0,000)
Urbana	0,205 (0,000)	0,170 (0,000)	0,223 (0,000)	0,242 (0,000)	0,255 (0,000)	0,269 (0,000)
Metropolitana	0,074 (0,000)	0,054 (0,015)	0,078 (0,000)	0,077 (0,000)	0,108 (0,000)	0,061 (0,000)
Norte	0,248 (0,000)	0,185 (0,000)	0,203 (0,000)	0,220 (0,000)	0,305 (0,000)	0,269 (0,059)
Sudeste	0,398 (0,000)	0,460 (0,000)	0,424 (0,000)	0,393 (0,000)	0,342 (0,000)	0,249 (0,000)
Sul	0,302 (0,000)	0,354 (0,000)	0,329 (0,000)	0,257 (0,000)	0,224 (0,000)	0,164 (0,001)
Centro-Oeste	0,294 (0,000)	0,318 (0,000)	0,300 (0,000)	0,316 (0,000)	0,327 (0,000)	0,324 (0,000)
Idade teve 1º filho	-0,028 (0,010)	-0,013 (0,000)	-0,010 (0,002)	-0,015 (0,000)	-0,012 (0,000)	-0,015 (0,003)
Educação = 1	-0,018 (0,695)	0,079 (0,295)	-0,002 (0,976)	-0,043 (0,494)	-0,059 (0,406)	0,030 (0,768)
Educação = 2	0,018 (0,612)	0,053 (0,258)	0,027 (0,542)	0,009 (0,817)	-0,059 (0,198)	-0,027 (0,715)
Educação = 3	0,052 (0,134)	0,095 (0,037)	0,076 (0,070)	0,081 (0,053)	0,133 (0,014)	0,187 (0,003)
Educação = 4	0,111 (0,002)	0,171 (0,000)	0,139 (0,000)	0,142 (0,000)	0,157 (0,000)	0,188 (0,001)
Educação = 5	0,203 (0,000)	0,244 (0,000)	0,247 (0,000)	0,248 (0,000)	0,209 (0,000)	0,285 (0,000)
Educação = 6	0,223 (0,000)	0,264 (0,000)	0,256 (0,000)	0,305 (0,000)	0,307 (0,000)	0,471 (0,000)
Educação = 7	0,313 (0,000)	0,327 (0,000)	0,320 (0,000)	0,393 (0,000)	0,394 (0,000)	0,542 (0,000)
Educação = 8	0,459 (0,000)	0,479 (0,000)	0,473 (0,000)	0,558 (0,000)	0,551 (0,000)	0,774 (0,000)
Educação = 9	0,451 (0,000)	0,500 (0,000)	0,474 (0,000)	0,542 (0,000)	0,593 (0,000)	0,683 (0,000)
Educação = 10	0,578 (0,000)	0,629 (0,000)	0,562 (0,000)	0,649 (0,000)	0,772 (0,000)	0,872 (0,000)
Educação = 11	0,858 (0,000)	0,803 (0,000)	0,837 (0,000)	0,984 (0,000)	1,104 (0,000)	1,196 (0,000)
Educação = 12	1,132 (0,000)	1,030 (0,000)	1,062 (0,000)	1,200 (0,000)	1,258 (0,000)	1,498 (0,000)
Educação = 13	1,454 (0,000)	1,387 (0,000)	1,444 (0,000)	1,478 (0,000)	1,763 (0,000)	1,871 (0,000)
Educação = 14	1,340 (0,000)	1,259 (0,000)	1,299 (0,000)	1,602 (0,000)	1,575 (0,000)	1,580 (0,000)
Educação = 15+	1,788 (0,000)	1,610 (0,000)	1,729 (0,000)	1,901 (0,000)	1,935 (0,000)	2,050 (0,000)

Nota: Valores-p entre parênteses.

tivo é de maior magnitude no quantil 0,1 (famílias de baixa renda), decrescendo (em valor absoluto) quando nos movemos para a cauda direita da distribuição condicional da renda, chegando a seu valor mínimo no quantil 0,75 (-0,142) e voltando a reverter essa tendência no quantil 0,9 (famílias de mais alto poder aquisitivo). Novamente, verificamos maior queda na renda familiar, em torno de 18%, nos quantis extremos (0,1 e 0,9). É interessante notar que as estimativas por EQT são de maior magnitude (em termos absolutos) do que as de regressão quantílica básica. Ou seja, o viés, quando não se corrige o problema da endogeneidade (usando-se o modelo tradicional de RQ), subestima

o impacto negativo da criação de filhos na renda familiar.²² Esses resultados podem estar refletindo as grandes desigualdades de renda no Brasil e significativas diferenças na taxa de fecundidade entre famílias pobres e ricas. Na nossa amostra, um pouco mais de 18% das famílias que têm pelo menos três filhos auferem rendimentos inferiores ao quantil 0,1 da distribuição de renda familiar e esse percentual sobe para aproximadamente 44% quando se consideram as famílias com renda inferior ao quantil 0,2. Enquanto apenas 7% das famílias pertencem à categoria das mais ricas (cuja renda familiar excede o quantil 0,9 da distribuição de renda) que têm três ou mais filhos. Isso demonstra que quase a metade das famílias mais numerosas (três ou mais filhos), encontra-se na cauda esquerda da distribuição de renda familiar de nossa amostra. A alta taxa de fecundidade e os baixos rendimentos das famílias pobres podem obrigar as mães a se ocuparem mais com a criação dos filhos, o que tenderia a afetar a oferta de trabalho das mesmas e a renda familiar. Por sua vez, o resultado para as famílias mais abastadas pode ser uma consequência de as mesmas optarem por investir mais na qualidade da criação dos filhos, como resultado de maior dedicação de seu tempo a eles.²³

Relativamente às *dummies* educacionais, a interpretação dos seus coeficientes deve ser vista como o efeito da educação da mãe na renda familiar. Vemos que esse efeito é mais elevado para aquelas que possuem maior nível educacional (em todos os quantis). Na mediana, por exemplo, a estimativa é de 0,142 para as mulheres que terminaram o primário, elevando-se para 1,901 no caso daquelas com 15 anos ou mais de estudo, o que demonstra a importância da educação na determinação dos rendimentos. Ainda no que concerne à educação, verificam-se dois outros resultados interessantes, que são detalhados a seguir. Para as mulheres com 11 ou mais anos de estudo, a estimativa do coeficiente é maior quando nos movemos em direção à cauda direita da distribuição de renda. Por exemplo, condicional a 15 anos de estudo, a estimativa do coeficiente é 1,61 para o quantil 0,1 e ela se eleva continuamente ao longo dos quartis, passando a atingir 2,05 no quantil 0,9. Esta tendência crescente ao longo dos quantis não se verifica para as mulheres de menor nível educacional.

22. Essa direção do viés também é verificada nos resultados para os Estados Unidos [Abadie, Angrist e Imbens (1998)].

23. Na versão anterior deste artigo, a variável dependente (renda familiar) incluiu a renda dos filhos (tendo sido a mesma excluída na corrente versão), então os resultados anteriores apontaram para um efeito positivo da criação de um terceiro ou mais filhos no quantil 0,1, sendo uma consequência da realidade brasileira na qual os filhos das famílias mais pobres tendem a trabalhar e têm considerável participação na renda familiar. Na primeira versão os resultados também apontaram para um efeito negativo no quantil 0,9, de maior magnitude que o encontrado na versão atual. Seguindo-se um raciocínio semelhante, os filhos das famílias de maior poder aquisitivo permanecem na escola e tendem a adiar o ingresso no mercado de trabalho, o que torna a renda familiar dessas famílias relativamente menor que a das demais (quando se inclui a renda dos filhos) na versão anterior. É importante também mencionar que foi estimado um modelo no qual a variável instrumental indicava gêmeos na primeira gravidez, seguindo a abordagem do artigo de Angrist e Evans (1998). Para o caso brasileiro, a amostra das famílias que tiveram gêmeos na primeira gravidez (e que preenchem os demais requisitos) é muito pequena, o que impossibilitou estimar com precisão o efeito da criação de filhos na distribuição de renda das famílias.

Em particular, condicional a se ter quatro anos completos de estudo (primário), no quantil 0,1 o valor da estimativa do parâmetro é 0,171, enquanto na mediana é de 0,142. Outro ponto importante a se notar é a maior variabilidade das estimativas condicional a um nível mais elevado de educação. Isso quer dizer que para pessoas mais qualificadas, condicional a um nível educacional em particular, a diferença entre os rendimentos familiares mais baixos e os mais altos é bastante significativa. Para as pessoas com 15 anos ou mais de estudos, conforme anteriormente especificado, essa diferença entre os quantis 0,1 e 0,9 é de 44%. Para as mulheres menos qualificadas, por sua vez, há menor variabilidade dos rendimentos familiares condicional ao nível educacional. Por exemplo, condicional a cinco anos de estudo, a diferença das estimativas dos coeficientes entre os quantis 0,9 e 0,1 é de 4,1%, e condicional a quatro anos de estudo essa diferença é de apenas 1,7%. Ou seja, os rendimentos familiares para as mães menos qualificadas são mais homogêneos, e para as que detêm maior qualificação, mais diferenciados. Os sinais para as demais variáveis foram como esperados: *a*) a renda familiar é maior para as mulheres de cor branca e para as famílias que habitam em áreas urbanas e metropolitanas; *b*) habitar na região Nordeste (*dummy* excluída) implica ter menores rendas familiares do que nas demais regiões (mantidas as demais características constantes). Outro resultado interessante é que para essas demais variáveis as estimativas dos coeficientes são bem semelhantes, havendo pouca variabilidade nas suas magnitudes ao longo da distribuição da renda.

5 CONCLUSÕES

Este artigo analisa o efeito da criação de um terceiro ou mais filhos nos quantis condicionais do log da renda familiar. Para tal, foi usado o estimador do EQT, desenvolvido recentemente por Abadie, Angrist e Imbens (1998 e 2002). O que é interessante na metodologia é que ela lida com o problema das *dummies* endógenas, tão comuns em economia (por exemplo, sindicatos, treinamentos), que possibilita calcular o efeito do tratamento para um subgrupo da população cujo *status* de tratamento é afetado pelo experimento induzido pelo instrumento (os *compliers*). Outra vantagem desse instrumental é que possibilita estimar o efeito de um tratamento (intervenção) em qualquer ponto (quantil) da distribuição condicional da variável dependente.

As estimativas do EQT indicam interessantes e importantes diferenças do efeito da criação de um terceiro ou mais filhos nos diferentes quantis condicionais estudados.²⁴ As estimativas encontradas variam no intervalo [0,142; 0,181]. Para o quantil 0,1 tem-se um efeito negativo em torno de 18%, que decresce

24. Esses resultados devem ser interpretados tendo em vista serem condicionais a *X*, *D* e *compliers*.

(em valor absoluto) à medida que nos movemos para os quantis mais elevados, atingindo o percentual de 14% no quantil 0,75. No quantil 0,9 essa tendência de queda se reverte e tem-se um efeito em torno de 18% nesse quantil. Dessa forma, os resultados apontam para uma queda maior na renda familiar nas caudas da distribuição, isto é, para as famílias mais pobres e as mais ricas. No entanto, esse efeito pode decorrer de diferentes razões, entre esses subgrupos extremos da população. A intuição nos leva a acreditar que a maior magnitude do efeito para as famílias mais pobres provém da necessidade de cuidar dos filhos, implicando uma redução das horas de trabalho ou mesmo uma permanência fora do mercado de trabalho. Para as famílias mais ricas, por sua vez, o resultado pode indicar uma escolha das mães no sentido de dedicarem maior parte de tempo à criação dos filhos, para assim elevarem o capital neles investido.²⁵ O mesmo padrão se verifica para as estimativas geradas pelo método de regressão quantílica tradicional (sem ponderação). No entanto, essas estimativas subestimam o impacto da criação de um terceiro ou mais filhos na renda familiar em todos os quantis estudados, sendo o viés resultante de não se estar corrigindo nesse modelo o problema da endogeneidade. Em termos gerais, tem-se que a criação de um terceiro ou mais filhos implica queda na renda familiar, independentemente do nível de renda da família.

ABSTRACT

Research on the consequences of childbearing on income and labor-supply is complicated by the fact that fertility is an endogenous variable, which may bias the conventional estimator. In this study we use the Quantile Treatment Effect (QTE) approach recently proposed by Abadie, Angrist e Imbens (1998 and 2002) to estimate that effect more adequately. It uses the parental preferences for a mixed sibling sex composition to construct an instrumental variable estimator, allowing it to vary along the quantiles of the family income distribution. The QTE estimator can be used to determine how an intervention affects the distribution of the response variable for individuals whose treatment status is changed by a binary instrument. It also has the advantage of collapsing to the conventional quantile regression when the treatment is exogenous. The QTE estimates with PNAD data for Brazil reveal that there are important differences, between the several quantiles, in the effect in family income of childbearing. It is a reduction that varies from 14% to 18% throughout the quantiles, and is larger in those at the extremes of the income distribution, in particular, for those that correspond to the families with smaller income.

25. O padrão brasileiro difere em parte dos resultados encontrados para os Estados Unidos [Abadie, Angrist e Imbens (1998)]. Esses autores encontram que as famílias mais afetadas são as de menor poder aquisitivo nos Estados Unidos, enquanto aquelas mais abastadas naquele país sofrem um pequeno impacto pela criação de um terceiro ou mais filhos, esse resultado é contrário ao encontrado no caso brasileiro. Essas diferenças talvez resultem de fatores culturais, de estilo de vida e estruturais entre os dois países. Pode acontecer de as mulheres americanas priorizarem mais que as brasileiras suas carreiras profissionais, relativamente ao investimento na qualidade da criação dos filhos, ou ainda, talvez haja maiores facilidades nos Estados Unidos que permitam maior eficiência na criação dos filhos, possibilitando que as mulheres tenham mais tempo para o trabalho além das horas dedicadas à família.

BIBLIOGRAFIA

- ABADIE, A. *Identification of treatment effects in models with covariates*. MIT, Department of Economics, 1997, mimeo.
- ABADIE, A., ANGRIST, J. D., IMBENS, G. W. *Instrumental variables estimation of quantile treatment effects*. NBER, 1998 (Working Paper, 229, 1-28).
- _____. *Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings*. Department of Economics, Harvard University, 2000 (Working Paper).
- _____. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, v. 70, p. 91-117, 2002.
- ANGRIST, J. D., EVANS, W. N. Children and their parent's labor supply: evidence from exogenous variation in family size. *American Economic Review*, v. 88, n. 3, p. 450-477, 1998.
- ANGRIST, J. D., IMBENS, G. W., RUBIN, D. B. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, v. 91, p. 444-472, 1996.
- BECKER, G. S. *Family: in the new palgrave, a dictionary of economics*. London: Macmillan, p. 281-285, 1987.
- BECKER, G. S., LEWIS, G. H. On the interaction between the quantity and quality of children. *Journal of Political Economy*, v. 81, p. 279-288, 1973
- BEN-PORATH, Y. Labor-force participation rates and the supply of labor. *Journal of Political Economy*, v. 81, p. 697-703, 1973.
- BEN-PORATH, Y., WELCH, F. Do sex preferences really matter? *Quarterly Journal of Economics*, v. 90, p. 285-307, 1976.
- BROWNING, M. Children and household economic behavior. *Journal of Economic Literature*, v. 30, p. 1.434-1.475, 1992.
- BUCHINSKY, M. Changes in the US wage structure 1963-87: application of quantile regression. *Econometrica*, v. 62, p. 405-458, 1994.
- CAMPÊLO, A. K., SILVA, E. N. Children and family income: instrumental variables estimation of quantile treatment effects. *Anais do XXIV Encontro da Sociedade Brasileira de Econometria*, 2002.
- CHAMBERLAIN, G. Quantile regression, censoring, and the structure of wages. In: SIMS, C. A. (ed.). *Advances in econometrics sixth world Congress 1*. Cambridge: Cambridge University Press, 1991.
- COLEMAN, M. T., PENCAVEL, J. Trends in market behavior of women since 1940. *Industrial and Labor Relations Review*, v. 46, p. 653-676, 1993.
- DE TRAY, D. Child quality and the demand for children. *Journal of Political Economy*, v. 81, p. 70-98, 1973.
- GOLDIN, C. *Career and family: college women look to the past*. NBER, 1995 (Working Paper, 5.188, 1-50).
- GONZAGA, G., SOARES, R. R. Determinação de salários no Brasil: dualidade ou não-linearidade no retorno à educação. *Revista de Econometria*, v. 19, p. 367-404, 1999.

- GRONAU, R. The effect of children on the housewife's value of time. *Journal of Political Economy*, v. 81, p. S169-S199, 1973.
- _____. Leisure, home production and work — the theory of allocation of time revisited. *Journal of Political Economy*, v. 84, p. 1.099-1.124, 1977.
- _____. Sex-related wage differentials and women's interrupted careers — the chicken or the egg. *Journal of Labor Economics*, v. 6, p. 277-301, 1988.
- HECKMAN, J. J., MACURDY, T. E. A life-cycle model of female labor supply. *Review of Economic Studies*, v. 47, p. 47-74, 1980.
- IMBENS, G. W., ANGRIST, J. D. Identification and estimation of local average treatment effects. *Econometrica*, v. 62, p. 467-476, 1994.
- KOENKER, R., BASSETT, G. Regression quantiles. *Econometrica*, v. 46, p. 33-50, 1978.
- KORENMAN, S., NEUMARK, D. Marriage, motherhood and wages. *Journal of Human Resources*, v. 27, p. 233-255, 1992.
- LEIBENSTEIN, H. *Economic backwardness and economic growth: studies in the theory of economic development*. New York: Wiley, 1957.
- LEWIS, H. G. *Union relative wage effects: a survey*. Chicago: University of Chicago Press, 1986.
- POTERBA, M. J., RUEBEN, K. S. *The distribution of public sector wage premia: new evidence using quantile regression methods*. NBER, 1994 (Working Paper, 4.734).
- RUBIN, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, v. 66, p. 688-701, 1974.
- _____. Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, v. 2, p. 1-26, 1977.
- _____. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*, v. 6, p. 34-58, 1978.
- VELOSO, F. A. Income composition, endogenous fertility and schooling investments in children. *Anais do XXII Encontro de Econometria*, 2000.
- WILLIS, R. J. A new approach to the economic theory of fertility behavior. *Journal of Political Economy*, v. 81, p. 514-564, 1973.
- _____. What have we learned from the economics of the family? *The American Economic Review*, v. 77, p. 68-81, 1987.

(Originais recebidos em setembro de 2005. Revistos em novembro de 2005.)