UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA – PIMES
MESTRADO EM ECONOMIA

EFEITO DO NÚMERO DE FILHOS NA DISTRIBUIÇÃO CONDICIONAL DA RENDA FAMILIAR: Uma Aplicação de Variáveis Instrumentais para Estimar o Efeito Quantílico de um Tratamento

MESTRANDO: EVERTON NUNES DA SILVA
ORIENTADORA: ANA KATARINA CAMPÊLO

Recife, maio de 2003.

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA – PIMES
MESTRADO EM ECONOMIA

EFEITO DO NÚMERO DE FILHOS NA DISTRIBUIÇÃO CONDICIONAL DA RENDA FAMILIAR: Uma Aplicação de Variáveis Instrumentais para Estimar o Efeito Quantílico de um Tratamento

Dissertação apresentada ao Programa de pós-graduação em Economia – PIMES – da Universidade Federal de Pernambuco como requisito para obtenção do título de Mestre em Economia

Mestrado: Everton Nunes da Silva
Orientadora: Ana KatarinaCampêlo

Recife, maio de 2003.

EVERTON NUNES DA SILVA

EFEITO DO NÚMERO DE FILHOS NA DISTRIBUIÇÃO CONDICIONAL DA RENDA FAMILIAR: Uma Aplicação de Variáveis Instrumentais para Estimar o Efeito Quantílico de um Tratamento.

Data da defesa: 13 de Junho de 2003

Banca Examinadora:

_____
Ana Katarina de Novaes Campelo
Orientadora

_____
Tarcísio Patrício de Araújo
Examinador interno

_____
Antonio Lisboa Teles da Rosa
Examinador externo - CAEN/UFC

À memória de meu pai, pela pessoa digna que foi.
À minha mãe, pelo amor incondicional que me foi dado.

- SUMÁRIO -

AGRADECIMENTOS

À minha orientadora Professora Ana Katarina Campêlo por sua contribuição ímpar no desenvolvimento desta dissertação, assim como pelos conselhos e incentivos no que concerne à pesquisa.

À minha família pela confiança, pelo carinho e pela compreensão nos momentos difíceis. Em especial à minha mãe pelos sábios conselhos e à minha irmã pelo apoio em todos os momentos.

Aos professores do Programa de Pós-graduação em Economia – PIMES – pela competência na transmissão de ensinamentos que muito acrescentou na minha formação acadêmica.

Aos colegas do mestrado e doutorado, em especial à Beatriz Mesquita, Anderson Saito, Pablo Camacho, Jorge Albino e Sylvan Reis, pelo convívio acadêmico e pela amizade formada.

À Patrícia Santana, Giliene Monteiro e Patrícia Alves pela competência em atender as minhas solicitações.

Ao professor Nilton Boscacci pelo constante estimulo ao que concerne à busca do conhecimento.

Ao Rômulo Rufino, Myrna Moura e Luan Moura Rufino por terem me dado o privilégio de conviver dois anos em sua família.

Ao Cícero Emanoel, André e Michely Moura pelo companheirismo.

Ao Erivan Alves pela amizade e pelo auxílio na área computacional.

À CAPES pelo suporte financeiro.

- RESUMO –

Esta dissertação visa investigar o impacto da criação de um terceiro filho na renda dos pais, que será denominada como renda familiar nesta dissertação. O impacto em questão é estimado em vários pontos da distribuição condicional da variável resposta (rendimento da família) por meio de uma variante da regressão quantílica denominada "efeito quantílico de um tratamento" (Abadie et al, 1998 e 2002). A grande maioria dos trabalhos se detém sobre o efeito médio de um tratamento, enquanto que o método utilizado neste estudo nos dá uma visão mais completa, posto que o mesmo estima este efeito na distribuição condicional da variável resposta e não apenas na sua média condicional. Com este método são estimadas regressões correspondentes ao diversos quantis condicionais de interesse da variável dependente, ao invés de uma única regressão para a média. Assim, se existirem assimetrias na relação entre a variável dependente e covariáveis, estas serão captadas. A estimação dos diversos quantis também nos dá uma idéia da forma da distribuição condicional da variável resposta, o que não é possível de se obter com a estimação de apenas uma curva, a média condicional, no caso da regressão de mínimos quadrados ordinários. Outra questão importante abordada neste trabalho refere-se a endogeneidade presente na relação entre o número de filhos (fecundidade) e os rendimentos da família. Grupos de controle e tratamento, bem como variáveis instrumentais serão utilizados para estimar a relação de interesse. Os resultados deste estudo mostraram que a criação de um terceiro filho causa um impacto negativo sobre a renda familiar em todos os quantis condicionais que varia entre 13,56% e 20,22% nos quantis estudados, sendo estes marginalmente maiores nos quantis inferiores, famílias de baixa renda.

Palavras-chave: renda, número de filhos, fecundidade, efeito de tratamento quantílico, variáveis instrumentais, regressão quantílica.

ABSTRACT

The purpose of this dissertation is to quantify the impact of raising a third child on parents income, which will be denoted by family income along this dissertation. In other words, it tries to analyze the effect of raising a third child in the conditional distribution of family income. This effect will be estimated at various points of the conditional distribution of the response variable (family income) using a recently proposed technique called ``quantile treatment effect'', a variant of quantile regression (Abadie et al, 1998 and 2002). The majority of studies in this area try to quantify the average effect by estimating the conditional mean. This study thus gives a more general picture of the relationship between fertility and family income by analyzing the effect on the conditional distribution of the dependent variable. This method allows one to estimate various regressions corresponding to the conditional quantiles of interest instead of a unique regression for the mean. Thus giving a more complete picture of the shape of the conditional distribution of the response variable, capturing the possible assimetries in the relationship of the dependent variable and its covariates. The econometric method used in this dissertation also deals with a well known problem in the relationship between fertility and income: endogeneity. It uses instrumental variable with treatment/control groups to estimate the relationship of interest. The results in this study show that a third child impacts negatively in parents income, the effect being marginally stronger in the lower (left) tails of the conditional distribution. The estimated quantiles were 0.1, 0.25, 0.5, 0.75 and 0.9, and the effect being in the range –0.1356 and –0.2022.

Key-word: family income, fertility, quantile treatment effect, instrumental variables, quantile regression, treatment/control groups.

- CAPÍTULO I -
Introdução

A decisão de ter filho(s) tem sido uma questão que tem gerado bastante interesse por parte de pesquisadores a partir da segunda metade do século passado, pelas implicações sócio-econômicas e demográficas que a mesma ocasiona. Com os trabalhos de Harvey Leibenstein (1957) e Gary Becker (1960), originou-se uma nova corrente em economia que mais tarde foi denominada de "Teoria da Família". Outros importantes autores desta corrente são: Becker & Lewis (1973), Ben-Porath (1973), Heckman (1980), De Tray (1973), Gronau (1973, 1977, 1988), Goldin (1995), Willis (1973), Schultz (1973), entre outros. Uma das principais "reivindicações" dessa corrente está relacionada ao fato de que pertence à família a essencial decisão de gerar e criar os filhos. Seguindo o raciocínio dessa abordagem, uma vez que se opta por ter filho(s), há, implicitamente, um comprometimento moral por parte dos pais de dedicarem uma determinada quantia de tempo e dinheiro para a geração e criação do(s) mesmo(s), o que representa uma mudança em seus hábitos que pode provocar alterações na oferta de trabalho dos pais. Neste sentido, caso ocorram tais modificações, estas podem provocar diferenciais nos rendimentos das famílias que possuem filho(s) e as que não o(s) possuem. O tamanho da família pode ainda ser um agravante destes diferenciais. Neste sentido, tornam-se importantes os estudos que visem uma melhor compreensão da relação entre criação de filhos e oferta de trabalho dos pais, sejam por razões teóricas, práticas ou para fins de políticas públicas.

A literatura afim defende que mães podem encontrar dificuldades em conciliar criação dos filhos e carreira profissional (ver 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 resultados apontam para uma relação entre salários mais baixos para as mulheres como conseqüência das saídas da força de trabalho para criar filhos (Gronau, 1988; Korenman e Neumark, 1992).

Segundo Willis (1987), a principal contribuição de Becker (1960) refere-se à sua hipótese de que existe uma parcela do gasto com os filhos que é endógena, devido ao fato de os pais receberem utilidade tanto pela quantidade de filhos, quanto pelo capital humano investido (CHI) nos mesmos[1]. Assim os pais devem decidir entre aumentar o tamanho da família ou intensificar o investimento em capital humano aos filhos existentes.

Segundo dados do IBGE, há uma nítida tendência de queda da taxa de natalidade, o que vem sinalizar uma opção dos pais em ter famílias menores. De 1980 a 2001, esta caiu de 32 para 20 nascimentos por mil habitantes, sendo que pelas projeções da mesma instituição, a trajetória de constante decréscimo se manterá, chegando a 14 nascimentos por mil habitantes em 2050. Uma possível explicação para este fenômeno pode estar relacionada ao custo de oportunidade de ter filho(s), o qual é crescente, principalmente nas grandes cidades. O aumento do "custo" relativo na geração e criação de filhos reflete a ascendente valorização do tempo dos pais[2] (ver

[1] 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 para mais unidades. A mesma lógica se aplica ao aumento no número de filhos. Assim, o custo marginal não é constante.

[2] A educação, como apontado pela literatura, é um importante fator que vem contribuindo para agregar valor ao trabalho. Deste modo, pode-se esperar que haja uma relação direta entre valorização do tempo dos indivíduos e escolaridade. Neste sentido, pode-se supor que o custo de oportunidade dos pais brasileiros (casal) aumentou, pois, segundo Maciel (2001), para a população brasileira de idade entre 18 e 65 anos de ambos os gêneros no período de 1992 a 1999, houve um aumento nos anos de estudo

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. O mesmo conclui 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 *per filho*, ao invés de aumentar o número de filhos.

Diante deste cenário, quantificar o impacto do tamanho da família sobre a renda familiar torna-se uma questão essencial tanto para os indivíduos (pais), que terão mais informação para tomar suas decisões acerca do tamanho ótimo da família, quanto para determinados órgãos governamentais, pois esta informação pode auxiliar na formação de políticas mais consistentes no que se refere à renda familiar e fecundidade, estando também estas questões relacionadas ao crescimento econômico de um país.

O objetivo desta dissertação é buscar maior entendimento da relação entre tamanho da família e renda familiar. Em particular, será estudado o efeito da criação de um terceiro filho na renda dos pais. Um terceiro filho pode ser visto como um número limite entre famílias grandes e pequenas. Para tal, usar-se-á a base de dados da Pesquisa Nacional de Amostragem Domiciliar (PNAD), coletada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), para o ano de 1999. Como os dados da PNAD fornecem características sobre a família e os indivíduos através de uma amostra representativa da população brasileira, os resultados e inferências deste estudo terão uma abrangência nacional.

A contribuição desta dissertação 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 uma análise

---

tanto para os homens quanto para as mulheres, sendo mais acentuado no caso das últimas (ver Maciel, 2001, página 3).

mais precisa e geral da relação de interesse. A maioria das aplicações em economia do trabalho e da família usa técnicas de estimação que possibilitam apenas análises mais simples, focando seus resultados sobre a média. No caso específico de um efeito de intervenção nos rendimentos da família, a análise do efeito médio torna-se pobre pelo caráter assimétrico da distribuição dos rendimentos. Esta assimetria é ainda mais proeminente no caso brasileiro devido à notória desigualdade de renda existente no país. Neste caso é fundamental o uso de técnicas econométricas que permitam o estudo deste efeito em diversos pontos da distribuição condicional dos rendimentos, dado que este efeito possivelmente é distinto para os diversos níveis de renda: os mais pobres (quantis inferiores ou cauda esquerda), de renda média (em torno da mediana) e mais ricos (quantis superiores ou cauda direita).A estimação dos diversos quantis nos dá, então, uma visão mais completa da relação entre número de filhos e renda familiar. A técnica que possibilita tal análise é a regressão quantílica, proposta por Koenker e Bassett (1978) e uma variante da mesma. Assim, será obtida uma estimativa da intervenção (no caso, o impacto de um terceiro filho) para cada quantil da distribuição condicional da renda familiar.

Outra contribuição importante deste estudo é o fato de o mesmo lidar com o problema da endogeneidade presente na relação entre fecundidade e renda. Ou seja, a teoria prevê que os determinantes do número de filhos e da renda familiar são determinados conjuntamente. Esta endogeneidade é 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 covariável. A presença de endogeneidade nesta 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 da regressão quantílica básica).

Uma possível solução para o problema da endogeneidade é utilizar variáveis instrumentais. Para tal, deve-se encontrar uma variável, denominada de *instrumento*, que seja correlacionada com a covariável e não-correlacionada com a variável resposta. Uma vez encontrado este instrumento, os parâmetros do modelo podem ser estimados consistentemente, pois a nova covariável (o instrumento) agora é exógena, isto é, determinada aleatoriamente relativamente à variavel resposta (para uma referência básica ver Greene, 2000). A grande questão é, então, encontrar um bom instrumento, o que, em algumas aplicações, não é uma tarefa fácil.

Nesta dissertação 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 tiverram dois filhos de gêneros distintos, um menino e uma menina. Sendo estes percentuais 53,01% e 46,70%, respectivamente, o que indica que famílias com os dois primeiros filhos de mesmo gênero têm uma chance maior de gerar um terceiro filho em torno de 13% relativamente às famílias que têm um menino e uma menina.

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: i) pelo próprio desejo em ter um menino e uma menina; e ii) 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 ambos os gêneros. Essa

hipótese foi provada empiricamente por eles, concluindo 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 de sexo, refere-se à tendência dos pais de querer no mínimo um filho do sexo masculino. Possíveis explicações para esse caso, segundo os autores[3], seriam que: i) meninos são esperados a contribuir mais do que meninas na renda familiar; e ii) meninos são uma fonte mais segura para dar suporte na velhice dos pais.

Dado o gênero do bebê ser 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 número de filhos e renda, dado ser correlacionado com o primeiro e independente do último[4].

As técnicas de regressão quantílica e variáveis instrumentais fazem parte do método econométrico utilizado nesta dissertação, que segue Abadie et al. (1998, 2002), como mencionado anteriormente. Estes autores desenvolveram uma variante da regressão quantílica com variáveis instrumentais, a qual foi denominada "Efeito Quantílico de um Tratamento''. Esta metodologia envolve ainda uma terceira técnica econométrica muito usada para quantificar o efeito de uma intervenção na variável de interesse, que se baseia em grupos de tratamento e controle. Esta intervenção é denominada na literatura como "tratamento'' e representada por uma variável dummy cujo valor 1 indica as pessoas que sofreram a intervenção (tratadas) ou que possuem a

[3] Essas possíveis diferenças nos custos ou benefícios associados a meninos e meninas são observadas, em maior grau, em países em desenvolvimento (Ben-Porath & Welch, 1976).
[4] Dentre os estudos que usaram tal instrumento estão os seguintes: Abadie et al. (1998) e Angrist e Evans (1998).

característica estudada, e o valor 0 corresponde às "não-tratadas" ou que não possuem a característica em questão. As primeiras são denominadas na literatura como *grupo de tratamento* e as últimas classificadas como *grupo de controle.* No presente estudo, a característica que representa o tratamento é a presença de um terceiro filho, então nossa dummy (endógena) é igual a 1 para as famílias que têm pelo menos 3 filhos e zero para aquelas que têm apenas dois filhos, o que vai nos possibilitar estudar o efeito da geração e criação de um terceiro filho.

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. Quanto maior a correlação entre estas dummies, melhor a qualidade do instrumento. A literatura (ver Angrist e Evans, 1998 e Abadie et al., 1998, 2002) denominou de "compliers" o subgrupo da população para o qual há perfeita correlação entre a dummy usada como instrumento e a dummy endógena. Para este subgrupo dizemos que o status de tratamento é afetado pelo experimento induzido pelo instrumento.

Em resumo, esta metodologia lida com o problema das dummies endógenas (técnica de variáveis instrumentais), possibilitando calcular o efeito de um tratamento (técnica de grupos de tratamento e controle) para um subgrupo da população (os compliers) em vários pontos da distribuição condicional da variável resposta (técnica de regressão quantílica). O desenvolvimento do processo para obtenção do efeito quantílico de tratamento para dummies endógenas será descrito em detalhes no próximo capítulo, sendo a tarefa mais difícil e inovadora do mesmo (proposto inicialmente por Abadie et al. 1998 e 2002) a identificação do subgrupo dos compliers. Note que este procedimento não visa extrair o subgrupo dos compliers da população, mas sim calcular o efeito do tratamento para este subgrupo usando a

totalidade dos dados. A identificação dos compliers será possível com o auxílio de pesos, cuja construção veremos adiante.

A dissertação como um todo é composta de cinco capítulos, incluindo esta introdução correspondente ao Capítulo I. O Capítulo II explana o referencial teórico que nos leva à identificação da subpopulação dos compliers e ao estimador do efeito médio local de tratamento; no Capítulo III é apresentada uma breve revisão acerca do método de regressão quantílica; enquanto que no Capítulo IV são descritos os dados, o modelo e os resultados da aplicação; e, finalmente, o Capítulo V sumariza as conclusões.

- CAPÍTULO II -

Identificação dos Compliers e o Efeito Médio Local de Tratamento (EMLT)

2.1 Introdução

O efeito quantílico de tratamento (EQT) para dummies endógenas (Abadie et al. 1998 e 2002) usado nesta dissertação é uma generalização do efeito médio local de tratamento (EMLT) (Angrist e Evans, 1998). O último concentra-se no efeito do tratamento na média condicional, enquanto que o primeiro é uma extensão para os quantis. Iremos, então, primeiramente descrever o processo a identificação do subgrupo dos compliers que foi proposto inicialmente por Angrist e Evans (1998) e que nos leva ao estimador do EMLT.

Em muitas aplicações econômicas (por exemplo, em economia do setor público e em economia do trabalho) busca-se estimar o efeito de uma intervenção ou de um tratamento em uma variável de interesse. A estimação deste efeito causal do tratamento nem sempre é uma tarefa fácil. Para que haja uma melhor compreensão desta questão, convém frisar que o efeito líquido de um tratamento é dado pela diferença entre o resultado *na presença* do tratamento e aquele *na ausência* do mesmo, para *um mesmo* indivíduo. Mas de fato não é possível observar *ambos* os resultados para *um mesmo* indivíduo.

Para ilustrar de forma mais intuitiva a estimação do efeito causal recorreremos a um exemplo ilustrativo. Suponha que pretendemos avaliar um programa social de

treinamento técnico em informática no intuito de verificar se o mesmo gera ou não um acréscimo no salário dos indivíduos que vão futuramente entrar no mercado de trabalho. Quando chegar o momento de observar o salário dos indivíduos, só teremos um resultado condicional: ou é um salário de um indivíduo que participou do treinamento ou um salário de um indivíduo que não participou do treinamento. Isto implica que, para *o mesmo* indivíduo, não observamos conjuntamente *seu salário com o treinamento (1)* e *seu salário sem o treinamento (2)*, ou seja, um indivíduo não pode estar simultâneamente no grupo de tratamento (tratados) e controle (não-tratados). Assim ficamos impossibilitados de conseguir isolar o efeito do tratamento, pois não é observada a situação contrafactual.

A técnica econométrica que usa grupos de tratamento e controle é um artifício utilizado na literatura para lidar com a dificulade descrita acima. A mesma faz uso de uma *proxy* que possibilita a estimação do resultado contrafactual não-observado. Aplicando esta técnica para o nosso exemplo, teríamos que o salário dos indivíduos do grupo de controle (não-treinados) seria usado como *proxy* para o salário dos indivíduos no grupo de tratamento (treinados) na situação (contrafactual) em que estes últimos não fossem tratados (ou não-treinados). A precisão desta técnica dependerá da qualidade da proxy, ou seja, de o grupo de controle ter características similares às do grupo de tratamento. Nestes tipos de análises é, então, fundamental a questão da classificação dos indivíduos e alocação dos mesmos nos grupos de tratamento e controle. Quando esta alocação não é aleatória, haverá um viés de seleção. Neste caso, diferenças na distribuição da variável resposta entre indivíduos tratados e não-tratados vão refletir não apenas o efeito causal do tratamento como também outros efeitos resultantes de características específicas de cada grupo[5].

[5] Outro exemplo pode ilustrar este ponto. Suponhamos um experimento clínico no qual o grupo tratado toma o remédio e o grupo de controle recebe um placebo. Vamos assumir ainda que, para este tipo de

Ainda, segundo Abadie (2001), estimadores do efeito de tratamento baseados na abordagem paramétrica tradicional podem ser fortemente viesados caso o termo referente ao erro estocástico não siga uma distribuição normal ou, de forma menos restrita, não satisfaça as condições: i) valor esperado do erro condicional às covariáveis ser zero: $E[\varepsilon \mid X] = 0$; e ii) homocedasticidade e ausência de autocorrelação: $E[\varepsilon\varepsilon' \mid X] = \sigma^2 I$. Desta forma, foram desenvolvidos na literatura métodos de estimação do efeito de tratamento baseados em técnicas não-paramétricas ou semi-paramétricas, no intuito de minimizar as consequências da não-observação dos resultados contrafactuais. O uso de tais técnicas econométricas que não requerem a especificação da distribuição do termo aleatório resulta em estimativas mais robustas do efeito de tratamento. A regressão quantílica classifica-se dentro da categoria de modelos semi-paramétricos, sendo, pois, um método mais robusto que o tradicional de regressão por mínimos quadrados.

Tendo em vista estas considerações, o método[6] usado na presente dissertação utiliza variáveis instrumentais para estimar o modelo de resposta média do tratamento com a adição de variáveis explicativas. Neste, a identificação do efeito médio do tratamento é obtido não-parametricamente, gerando estimadores mais robustos na ausência de normalidade. Segundo Spanos (1999), a abordagem não-paramétrica possui certas vantagens sobre a abordagem paramétrica, como o estabelecimento de um menor número de pressupostos, minimizando assim o problema de má-especificação do modelo.

---

problema de saúde em estudo, as crianças enfermas sejam biologicamente mais propensas a se curarem que os adultos acamados pela doença. Se o experimento não for aleatório e acontecer de uma quantidade significativa de crianças ser alocada para o grupo de tratamento, a diferença entre os percentuais de cura dos grupos de tratamento e controle refletirá não apenas o efeito do tratamento como também a propensão natural das crianças em atingir a cura, o que resultaria em um viés positivo. Concluí-se, então, que a precisão da estimativa do efeito do tratamento está fortemente atrelada à aleatoriedade do experimento.

[6] Ver Abadie, Angrist & Imbens (1998), Abadie (2001), Frölich (2002).

Dentre as vantagens da introdução de variáveis explicativas no modelo está a de possibilitar a avaliação da validade do instrumento[7], além de serem utilizadas para controlar diferentes atributos individuais observáveis.

A técnica de variáveis instrumentais, por sua vez, representa um poderoso instrumento no processo de identificação do efeito causal, o que lhe confere grande aplicabilidade em modelos que buscam estimar o efeito médio do tratamento (Heckman, 1995). As variáveis instrumentais atuam no sentido de lidar com a endogeneidade entre o tratamento, que denominaremos por D, e a variável resposta de interesse, Y, pelo uso de uma terceira variável, o instrumento (para o qual é usada a notação Z), a qual está correlacionada com D e não com Y. Deste modo, o instrumento atribui variação exógena à relação entre D e Y para todos os casos em que Z = D, isto é, os casos em que uma mudança em D é provocada por uma alteração exógena em Z. Na literatura, esta subpopulação, para a qual Z = D, tem sido denonata pelo termo *compliers,* enquanto que o termo Efeito Médio Local do Tratamento (EMLT) refere-se ao impacto gerado pelo efeito causal do tratamento sobre a variável resposta.

Este capítulo tem como finalidade mostrar como se estima o efeito médio local do tratamento e, para tal, compreende quatro seções, incluindo esta introdução. A Seção 2.2 expõe o problema da identificação do efeito causal, a Seção 2.3 introduz o problema de identificação do método de variáveis instrumentais (VI) e a Seção 2.4 identifica o efeito médio local do tratamento.

2.2 O problema da identificação

[7] Um bom instrumento deve ser independente não só da variável resposta como também das variáveis explicativas. Assim, com a introdução de variáveis explicativas no modelo, pode-se mensurar a

No modelo utilizado na presente dissertação, Y representa o log da renda familiar das famílias com dois ou mais filhos e D é o indicador de tratamento binário, que indica famílias com mais de dois filhos. Os resultados potenciais são definidos por $Y_{D,}$ que representa resultado potencial da variável resposta, Y, condiciontal ao status de tratamento, D. Nesta aplicação, Y_1 representa o log da renda familiar das famílias que tiveram o terceiro filho (D = 1) e Y_0 indica o log da renda familiar das famílias que não tiveram um terceiro filho (D = 0). O que se busca identificar neste caso é o efeito do tratamento sobre o log da renda familiar, o qual pode ser definido por $(Y_1 - Y_0)$. Entretanto, não é possível observar ambos Y_1 e Y_0 para o mesmo indivíduo, como discutido anteriormente. Considerando que sempre um dos resultados potenciais (a situação contrafactual) não é observado para uma mesma pessoa, não é possível computar o efeito causal do tratamento para o indivíduo. Surge, dessa forma, um problema de inferência causal, resultante do fato de ser necessária a comparação entre o resultado observado (caso com tratamento) e o não-observado (caso sem tratamento), para um mesmo indivíduo. Formalizando, tem-se:

$$E[Y_1 \mid D = 1] - E[Y_0 \mid D = 0] = \{E[Y_1 \mid D = 1] - E[Y_0 \mid D = 1]\} + \{E[Y_0 \mid D = 1] - E[Y_0 \mid D = 0]\} \quad (2.1)$$

O termo do lado esquerdo representa a diferença entre os resultados médios observados por status de tratamento, isto é, a diferença entre os resultados médios provindo do grupo de tratamento relativamente ao grupo de controle. O primeiro termo do lado direito da equação é o que se procura estimar: o efeito médio do

---

validade ou, em outros termos, quão "bom" é o instrumento.

tratamento sobre o log da renda familiar dos tratados[8]. O segundo termo do lado direito indica o viés ocasionado pela endogeneidade resultante da alocação do tratamento (Abadie, 2001). Mais explicitamente, o segundo termo mostra a diferença entre o resultado médio do grupo de controle e aquele que seria observado para o grupo dos tratados caso estes não tivessem recebido o tratamento. O viés surge quando a *proxy* usada, resultados do grupo de controle, não explica de forma adequada o valor contrafactual não-observado das pessoas no grupo de tratamento. No caso em que a alocação dos grupos de tratamento e controle é feita de forma aleatória, este viés desaparece, pois ambos os grupos neste caso são uma amostra da população, não apresentando assim características distintas e particulares que poderiam estar correlacionadas com a variável resposta (que se resumiria no problema da endogeneidade). O artigo de Imbens e Angrist (1994) sugere um procedimento que contorna o problema citado (viés causado pela não-aleatoreidade na alocação do tratamento) através do uso de uma variável instrumental que possibilita a identificação de efeitos causais para os compliers. Este procedimento será detalhado na próxima seção. Nesta dissertação usamos ainda uma metodologia mais atual, proposta por Abadie et al (1998, 2002), a qual é uma extensão do artigo de Imbens e Angrist (que calcula o efeito médio) para o modelo mais abrangente dos quantis condicionais.

2.3 Identificação por Variáveis Instrumentais (VI)

[8] Se a expectativa é um operador linear, esse termo pode ser escrito da seguinte forma: $E[Y_1 - Y_0 \mid D_1]$ (Abadie et al, 1998).

Em muitas aplicações, estimar a relação causal entre duas variáveis pode se tornar uma tarefa difícil, caso estas tenham uma relação endógena. Neste contexto, variáveis instrumentais podem ser usadas para criar uma variação exógena para a variável de tratamento para que, desta forma, possa ser estimado o efeito causal na variável resposta. No caso específico desta dissertação, que visa estimar o impacto do terceiro filho sobre o log da renda familiar, será usada uma variável instrumental baseada na composição do sexo dos dois primeiros filhos como um indicador para a geração do terceiro filho[9].

Com a introdução do instrumento, os efeitos causais de interesse são colocados em termos de resultados potenciais, descritos por Y_{zd}, ou seja, a renda familiar quando se tem $Z = z$ e $D = d$, enquanto D_z representa o indicador de mais de dois filhos quando temos $Z = z$. Assim definida, a variável dependente observada é:

$$Y = [Y_{00} + (Y_{01} - Y_{00}) \,.\, D_0] \,.\, (1 - Z) + [\, Y_{10} + (Y_{11} - Y_{10}) \,.\, D_1] \,.\, Z \qquad (2.2)$$

Por exemplo, quando se observa $Z = 1$ e $D = 1$, a variável resposta toma o valor Y_{11}, que significa o valor da renda familiar quando Z=1 e D=1, mais precisamente, famílias tiveram os dois primeiros filhos do mesmo sexo e, por esse motivo, tiveram o terceiro filho. Os demais resultados potenciais são definidos de forma similar.

Para que o modelo seja consistente, alguns pressupostos fazem-se fundamentais. São descritas abaixo as principais hipóteses necessárias para a identificação do efeito médio local do tratamento. Seja Z o instrumento binário, o qual indica famílias com os dois primeiros filhos do mesmo sexo, D a dummy de

[9] Na ocorrência de mesmo gênero para os dois primeiros filhos, os pais serão significativamente mais propensos a continuar ter filhos (ver Ben-Porath e Welch, 1976).

tratamento que indica famílias com mais de dois filhos, X a matriz *n x q* de variáveis explicativas e Y os resultados potenciais.

(i) **(Independência)** $\{Y_{11}, Y_{10}, Y_{01}, Y_{00}, D_1, D_0\}$ são conjuntamente independentes de Z dado X;

(ii) **(Exclusão)** $P(Y_{1D} = Y_{0D} |X) = 1$;

(iii) **(Distribuição não-degenerada)** $P(Z = 1 \mid X) \in (0,1)$;

(iv) **(Primeiro estágio)** $E[D_1 \mid X] \neq E[D_0 \mid X]$;

(v) **(Monotonicidade)** $P(D1 \geq D0 \mid X) = 1$.

Os pressupostos (i)[10] e (ii) referem-se às propriedades básicas que um instrumento deve possuir: (i) o instrumento deve ser independente dos resultados potenciais dado X e (ii) a única forma de Z influenciar Y é pelo seu efeito em D. A hipótese (iii) requer que a distribuição do instrumento não seja degenerada e (iv) diz que é necessário haver uma relação entre D e Z para que se possa realizar o primeiro estágio dentro da teoria de variáveis instrumentais. Por fim, o pressuposto (v), monotonicidade, diz que o instrumento só pode afetar a variável de tratamento, D, em uma única direção que, no caso desta aplicação, significa diz que a presença dos dois primeiros filhos de mesmo sexo (Z = 1) afetaria a dummy endógena apenas na direção de aumentar a probabilidade de ter o terceiro filho. Esta última hipótese é crucial para o processo de obtenção do EMLT, pois permite que a população possa ser dividida em três sub-populações: os *compliers*, os *alwaystakers* e os *nevertakers*, que serão descritas em maiores detalhes na próxima seção, a qual apresenta a construção teórica que busca identificar o EMLT.

2.4 Identificação do Efeito Médio Local do Tratamento (EMLT)

Com a utilização da técnica de variáveis instrumentais, pode-se resolver o problema de identificação do efeito causal de tratamento, mas apenas para o grupo de indivíduos (os compliers) cujo status de tratamento é afetado pelo instrumento. Para este grupo a ocorrência ou não de filhos do mesmo sexo (experimento induzido por Z) implica, respectivamente, na geração ou não de um terceiro filho (status do tratamento denotado por D), isto é, os compliers são o grupo de individuos para os quais Z=D. Mais precisamente, os compliers são aqueles indivíduos para os quais se Z=1 então D seria também 1 e se Z fosse igual a zero, teríamos D=0. Comumente, no entanto, há casos em que temos $Z \neq D$. Quando isto acontece dizemos que há imperfeita complacência (*imperfect compliance*). Neste caso surge uma dificuldade para identificar o efeito causal de tratamento, advinda do fato de que não é possível saber quem são os compliers. A impossibilidade de identificar os compliers acontece porque só é observado um resultado potencial para uma mesma família. Na nossa proposta de estudo, ou as famílias têm dois filhos que são do mesmo sexo (Z=1) ou de sexo oposto (Z=0). Não podemos ter *ambas* as situações (com seus respectivos resultados potenciais de D e Y), apenas uma delas se verifica. O artigo de Imbens e Angrist (1994) propõe uma forma de contornar este problema que é o principal desenvolvimento para chegarmos ao estimador do EMLT. O teorema a seguir (Imbens & Angrist, 1994) sintetiza este último resultado.

---

[10] O pressuposto (i) também é chamado de ignorabilidade de Z. Nas palavras de Abadie (2001, pág. 5): "Z is 'as good as randomly assigned' once we condition on X". Ou seja, condicional a X, o experimento induzido pelo instrumento é aleatório.

Teorema 1: Usando os pressupostos (i)-(v) e assumindo que as expectativas relevantes são finitas, temos:

$$\frac{E[Y \mid Z = 1, X] - E[Y \mid Z = 0, X]}{E[D \mid Z = 1, X] - E[D \mid Z = 0, X]} = E[Y_1 - Y_0 \mid X, D_1 > D_0] \quad (2.3)$$

O Teorema 1 nos dá o Efeito Médio Local do Tratamento (EMLT) para os compliers, o grupo de indivíduos para os quais $D_1 > D_0$[11]. Note que estes indivíduos não podem ser identificados, pois, para tal, seria necessário observar ambos os resultados para uma mesma pessoa (D_1 e D_0), o que não acontece na prática, ou é observado D_1 ou D_0 para um mesmo indivíduo. No entanto, é possível identificar certos indivíduos como "não-compliers" (que consistirão nos grupos de indivíduos denominados de "always-takers" e "never-takers"), o que permitirá o isolamento do grupo dos compliers através de exclusão, como será demonstrado a seguir. Este é o resultado mais importante desenvolvido no artigo do Imbens e Angrist (1994). A grande contribuição desta metodologia é permitir o isolamento do efeito do tratamento (para um grupo de indivíduos, os compliers) nos casos em que temos um experimento não-aleatório, ou aleatório com complacência parcial.

2.4.1 O Status de Tratamento é Ignorável para os Compliers

Como mencionado anteriormente, a importância de o experimento ser aleatório reside na garantia de os resultados potenciais serem independentes da *alocação* do

[11] Como descrito anteriormente, para os compliers temos D=1 se Z=1 e D=0 se Z=0, ou seja, a sub-população dos compliers corresponde ao grupo dos indivíduos que foram afetados pelo experimento induzido por Z.

tratamento[12]. Neste caso dizemos que a alocação do tratamento é *ignorável* e o efeito do tratamento pode ser computado pela diferença entre as distribuições da variável resposta por status de tratamento (Rubin, 1978). No caso desta dissertação, esta relação não é ignorável, dada a existência de endogeneidade entre fecundidade e renda. Apenas assume-se que o instrumento (Z), no exemplo o sexo dos dois primeiros filhos, é independente da renda potencial das famílias. O papel do instrumento é "fornecer" uma variação exógena ao tratamento para que desta forma a relação entre D e Y possa ser estimada de forma consistente. Contudo, permanece a possibilidade de haver o problema da complacência parcial. O estimador do efeito médio local do tratamento, desenvolvido no trabalho seminal de Imbens e Angrist (1994), contorna esta dificuldade através da estimação do efeito do tratamento para os compliers, grupo para o qual o experimento induzido pelo instrumento de fato afeta o status do tratamento. Para este grupo tem-se que Z = D, então é possível substituir a dummy de tratamento pelo instrumento, o que nos dá independência entre os resultados potenciais e o tratamento para este grupo particular dos compliers. Este resultado é formalizado no Lema abaixo:

Lema 1 : Assumindo os pressupostos (i)-(v) e condicionando a X, os status de tratamento (D) é ignorável para **compliers**: $(Y_1\ Y_0) \perp\!\!\!\perp D \mid X, D_1 > D_0$.

O resultado deste Lema advém da hipótese (i) assumida:$(Y_1\ Y_0, D_1, D_0) \perp\!\!\!\perp Z \mid X$, que nos dá $(Y_1\ Y_0) \perp\!\!\!\perp Z \mid X, D_1 = 1, D_0 = 0$. Esta última condição diz que os resultados potenciais $(Y_1\ Y_0)$ e instrumento (Z) são independentes. No caso em que $D_1 = 1$ e $D_0 = 0$ podemos substituir D por Z. Então, para este grupo em particular, os

[12] Quando o experimento não é aleatório surge o problema da endogeneidade, que resulta em

compliers, pode-se estimar o efeito do tratamento comparando as médias condicionais por status de tratamento, mesmo se a alocação do tratamento não for independente dos resultados potenciais. Este resultado é descrito na relação abaixo:

$$E [Y \mid D = 1, D_1 > D_0, X] - E [Y \mid D = 0, D_1 > D_0, X] = E [Y_1 - Y_0 \mid X, D_1 > D_0] \quad (2.4)$$

Vimos que não é possível identificarmos a população dos compliers, o que impossibilita a aplicação do resultado acima. Para que o mesmo possa ser operacionalizado na prática, define-se a seguinte função de D, Z e X:

$$k = k (D, Z, X) = 1 - \frac{D.(1 - Z)}{1 - E [Z \mid X]} - \frac{Z .(1 - D)}{E[Z \mid X]} \quad (2.5)$$

Note que k assume o valor de 1 se D = Z. Para os demais casos, k assume um valor negativo. Esta função "identifica" assim os *compliers,* grupo para o qual D=Z (Abadie, (1997). Neste caso, qualquer parâmetro definido como uma solução de uma condição de momento envolvendo (Y,D,X) pode ser identificado para os compliers como mostra o Lema 2 abaixo.

Lema 2. Seja ψ (Y, D, X) uma função real mensurável de (Y, D, X) e assumindo pressupostos (i)-(v),

$$\frac{E [k. \psi (Y, D, X)]}{P(D_1 > D_0)} = E [\psi (Y, D, X) \mid D_1 > D_0] \quad (2.7)$$

---

estimativas viesadas do efeito do tratamento.

Para um melhor entendimento deste resultado, definimos dois grupos de indivíduos: os *always-takers* e os *never-takers*. O primeiro refere-se aos indivíduos que sempre terão famílias grandes (acima de dois filhos), independentemente do instrumento ($D_1 = D_0 = 1$). O segundo compreende indivíduos que nunca terão famílias grandes, independentemente do instrumento ($D_1 = D_0 = 0$). Apoiado no pressuposto de monotonicidade, pode-se reescrever a função ψ em termos não só dos *compliers*, mas também dos *alwaystakers* e *nevertakers.*

$$\begin{aligned} E[\psi \mid X] = {} & E[\psi \mid X, D_1 > D_0] \,.\, P(D_1 > D_0 \mid X) \\ & + E[\psi \mid X, D_1 = D_0 = 1] \,.\, P(D_1 = D_0 = 1 \mid X) \\ & + E[\psi \mid X, D_1 = D_0 = 0] \,.\, P(D_1 = D_0 = 0 \mid X) \end{aligned} \tag{2.8}$$

Rearranjando os termos,

$$\begin{aligned} E[\psi \mid X, D_1 > D_0] = {} & \frac{1}{P(D1 > D_0 \mid X)} \{E[\psi \mid X] - E[\psi \mid X, D_1 = D_0 = 1].P(D_1 = D_0 = 1 \mid X) \\ & - E[\psi \mid X, D_1 = D_0 = 0] \,.\, P(D_1 = D_0 = 0 \mid X)\} \end{aligned} \tag{2.9}$$

Como a população de *compliers* é definida quando Z = D, por monotonicidade, pode-se identificar os demais grupos. Assim, *alwaystakers* são indivíduos com D = 1 e Z = 0, ao passo que *nevertakers* são os casos em que D = 0 e Z = 1. Nesses dois casos específicos, existe como conhecer cada um desses grupos, pois somente um evento é necessário para identificá-los, diferentemente dos *compliers* que requerem os dois eventos, (Z = 0, D = 0) e (Z = 1, D = 1), o que não pode ser identificado para o mesmo indivíduo. Pelo pressuposto de Z ser ignorável dado X, tem-se:

$$E[\psi \mid X, D_1 = D_0 = 1] = E[\psi \mid X, D = 1, Z = 0] \quad (2.10)$$

$$= \frac{1}{P(D = 1 \mid X, Z = 0)} E\left(\frac{D.(1 - Z).\psi}{P(Z = 0 \mid X)} \,\middle|\, X \right)$$

e

$$E[\psi \mid X, D_1 = D_0 = 0] = E[\psi \mid X, D = 0, Z = 1] \quad (2.11)$$

$$= \frac{1}{P(D = 0 \mid X, Z = 1)} E\left(\frac{(1 - D).Z\,\psi}{P(Z = 1 \mid X)} \,\middle|\, X \right)$$

De forma similar, pode-se encontrar a proporção de indivíduos *always-takers* e *never-takers*.

$$P(D_1 = D_0 = 1 \mid X) = P(D = 1 \mid X, Z = 0) \quad (2.12)$$

$$P(D_1 = D_0 = 0 \mid X) = P(D = 0 \mid X, Z = 1) \quad (2.13)$$

Identificada a proporção correspondente a cada grupo, segue-se para a substituição desses valores na equação referente ao valor esperado dos *compliers.* Tem-se, então:

$$E[\psi \mid X, D_1 > D_0 = 0] = \frac{1}{P(D_1 > D_0 \mid X)} E\left[\left(1 - \frac{D.(1 - Z)}{P(Z = 0 \mid X)} - \frac{(1 - D).Z}{P(Z = 1 \mid X)} \right) \psi \,\middle|\, X \right] \quad (2.14)$$

O passo final é aplicar o teorema de Bayes e integrar sobre X (Abadie et al., 1998). Com esse procedimento, obtém-se, então, o valor esperado da população para os *compliers*.

Neste capítulo expomos como se identifica o Efeito Médio Local do Tratamento (EMLT), seguindo a metodologia desenvolvida por Imbens e Angrist (1994). Contudo, Abadie et al (1998) estenderam o modelo para os quantis condicionais da distribuição da variável resposta, criando o Efeito Quantílico do Tratamento (EQT), o qual será utilizado nesta dissertação. Ao invés de obter o efeito do tratamento para o caso da média condicional, o EQT estima o efeito do tratamento aos diferentes quantis da distribuição condicional da variável resposta, captando, desta forma, uma melhor caracterização da relação entre Y e D. Assim, o próximo capítulo busca apresentar uma breve exposição da técnica de regressão quantílica, para que desta forma possamos mostrar como se estima o EQT (que será apresentado no Capítulo IV).

- CAPÍTULO III -

Regressão Quantílica

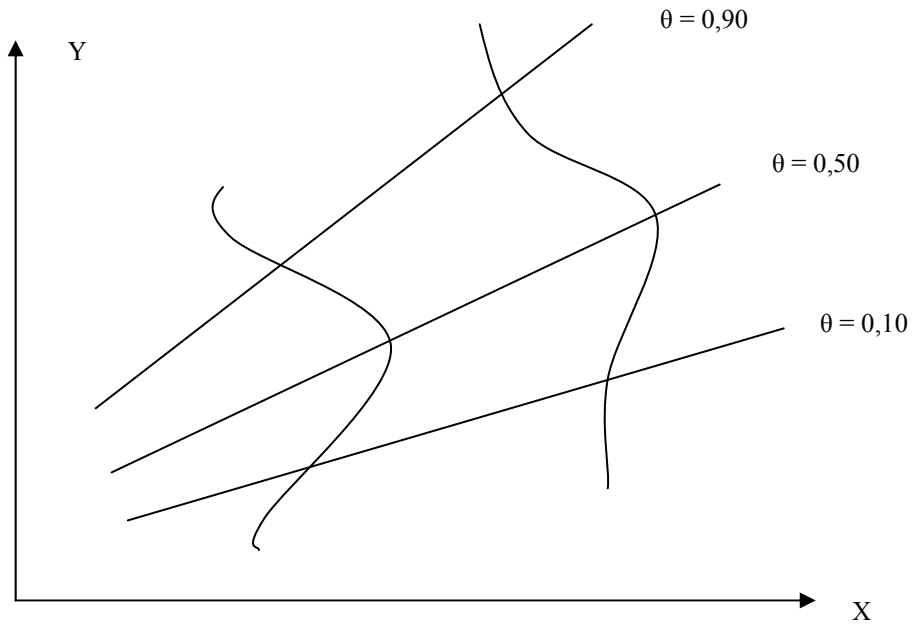
3.1 Introdução

Data de 1805, com o trabalho desenvolvido por Legendre, a primeira publicação sobre o que se convencionou chamar de Mínimos Quadrados Ordinários (MQO). Dos seus primórdios até os anos recentes, popularizou-se como uma ferramenta muito utilizada para estudar a relação entre variáveis, com o intuito de estimar e/ou prever a resposta média da variável dependente, Y, em termos dos valores assumidos pelas covariáveis, X. Como Y está sendo estimado com relação ao seu valor médio, assume-se, implicitamente, que sua flutuação se distribui igualmente em relação a X. Em outros termos, deve haver simetria na distribuição de Y dado X. Caso contrário, a média fornece uma visualização incompleta da distribuição do Y condicional a X.

Alguns fatos importantes ocasionaram o sucesso dos MQO, tais como: a facilidade no tratamento computacional; ii) ser um método que permite estimar a média condicional; e iii) possuir um estimador de mínima variância entre os estimadores não-viesados para os casos em que a função de distribuição da variável resposta é gaussiana (normal). Entretanto, como argumenta Koenker & Basset (1978), os MQO são extremamente sensíveis a valores extremos (outliers), como é o caso de distribuições não-gaussianas, produzindo estimadores com pouca precisão.

Uma visão mais completa pode ser obtida pelo método de regressão quantílica, pois este é uma técnica estatística que visa estimar e/ou inferir condicionalmente aos quantis da distribuição de Y. Desta forma, pode-se obter uma regressão para cada quantil ao invés de somente uma para a média, como é o caso de MQO. No gráfico 1,[13] onde a distribuição dos erros apresenta assimetria ou heteroscedasticidade, mostra que estimativas baseadas na média condicional não captam uma informação precisa da relação entre Y e X. Contudo, pelo método de regressão quantílica é possível obter um estimador robusto para cada quantil condicional, "mapeando" de forma mais completa as informações contidas na relação entre Y e X.

Gráfico 1. Regressão quantílica para os quantis 0,10; 0,50 e 0,90



Não há uma especificação rígida no que concerne ao número de quantis, podendo esta variar de pesquisa a pesquisa. Contudo, há uma tendência em dividir a população em cinco quantis, os quais são: 0,10; 0,25; 0,50; 0,75; e 0,90.

[13] Este gráfico é uma reprodução de um exemplo citado pelo Professor Belluzo Júnior no minicurso sobre Regressão Quantílica, realizado no XXIV Encontro Nacional da Sociedade Brasileira de Econometria, Rio de Janeiro, dezembro de 2002.

Neste capítulo é apresentada uma breve revisão acerca da técnica de regressão quantílica. Assim, o capítulo se divide em quatro seções, incluindo esta introdução. A Seção 3.2 sintetiza algumas vantagens de RQ sobre os MQO, a Seção 3.3 apresenta a técnica de regressão quantílica e a Seção 3.4 comenta algumas importantes propriedades de RQ.

3.2 Vantagens de Regressão Quantílica

Algumas vantagens inerentes à regressão quantílica sobre os MQO podem ser listadas da seguinte forma, como apontado por Ribeiro (1997):

- A técnica de regressão quantílica permite caracterizar toda distribuição condicional de uma variável resposta a partir de um conjunto de regressores.
- Regressão quantílica pode ser usada quando a distribuição não é gaussiana.
- Regressão quantílica é robusta a outliers.
- Por utilizar a distribuição condicional da variável resposta, podem-se estimar os intervalos de confiança dos parâmetros e do regressando diretamente dos quantis condicionais desejados.
- Dado os erros não possuírem uma distribuição normal, os estimadores provenientes da regressão quantílica podem ser mais eficientes que os estimadores por meio de MQO.

- Como a regressão quantílica pode ser representada como um modelo de programação linear facilita a estimação dos parâmetros. Muitos pacotes econométricos já possuem comandos próprios para esta finalidade, tais como S-PLUS, Stata, SHAZAM, entre outros.

3.3 A técnica de Regressão Quantílica

A técnica da regressão quantílica foi desenvolvida pelo trabalho seminal de Koenker & Bassett (1978), o qual deve ser visto como uma generalização do modelo de regressão de Mínimos Desvios Absolutos (MDA), L_1 ou regressão mediana para o caso do modelo de regressão linear, permitindo estimar não só a mediana, mas também outros quantis da distribuição de Y.

Antes de expor a técnica de regressão quantílica faz-se necessário apresentar a função quantil. Seja Y um vetor de variáveis aleatórias que assume valores reais caracterizado por sua função distribuição, dada por:

$$F(y) = \text{Prob}(Y \leq y) \tag{3.1}$$

e definindo θ entre (0, 1), obtém-se a função quantil:

$$F^{-1}(\theta) = Q(\theta) = \inf\{y : F(y) \geq \theta\}, \tag{3.2}$$

em que o θ é denominado o θ-ésimo quantil de Y, sendo $\theta = ½$ o quantil referente à mediana.

Segundo Maciel (2001), uma importante propriedade de função quantil concerne ao fato que para $-\infty \leq y \leq +\infty$ e $0 \leq \theta \leq 1$, $F(y) \geq \theta$ se e somente se $Q(\theta) \leq y$. Assim, tem-se Y identicamente distribuída a $Q(\theta)$.

Os parâmetros estimados pelo método de regressão quantílica são obtidos pela solução de um problema de minimização. Definindo a perda (erro) pela função $\rho_\theta(u) = u[\theta - I(u < 0)]$, tendo θ entre (0, 1), deve-se encontrar $\hat{y}$ que minimize o erro esperado. Para tal, minimiza-se a seguinte equação:

$$E\rho_\theta(Y - \hat{y}) = (\theta - 1)\int_{-\infty}^{\hat{y}}(y - \hat{y})dF(y) + \theta\int_{\hat{y}}^{\infty}(y - \hat{y})dF(y) \qquad (3.3)$$

Desde que F seja monotônica e diferenciando a equação com respeito a $\hat{y}$, tem-se algum elemento de $\{y : F(y) = \theta\}$, que minimiza o erro esperado. Para os casos em que há somente uma solução, $\hat{y} = F^{-1}(\theta)$. Casos contrários, há um intervalo de θ-ésimo quantis, dos quais deve-se escolher o melhor elemento dentre eles.

Substituindo a F pela função de distribuição empírica, chega-se ao seguinte problema de minimização:

$$\int \rho_\theta(y - \hat{y})dF_n(y) = n^{-1}\sum_{i=1}^{n}\rho_\theta(y_i - \hat{y}) = \min! \qquad (3.4)$$

ou reescrevendo o modelo na forma original (Koenker & Basset, 1978)

$$\min_{b \in \Re}\left\{\sum_{t \in \{t:y_t \geq b\}}\theta\left|y_t - b\right| + \sum_{t \in \{t:y_t < b\}}(1 - \theta)\left|y_t - b\right|\right\} \qquad (3.5)$$

chega-se ao θ-ésimo quantil amostral.

A regressão quantílica pode ser vista como uma extensão natural dos quantis amostrais para uma classe mais geral, na qual os quantis condicionais têm a forma linear. Assim, generalizando para o caso linear ($y_t = X_t\beta + \varepsilon_t$), em que a variável dependente, Y, é um vetor n *x* 1 de variáveis aleatórias independentes; X é uma matriz n *x* k de variáveis explicativas; β é um vetor k *x* 1 de coeficientes de regressão; e ε é um vetor n *x* 1 de erros, a função objetivo assume a seguinte forma:

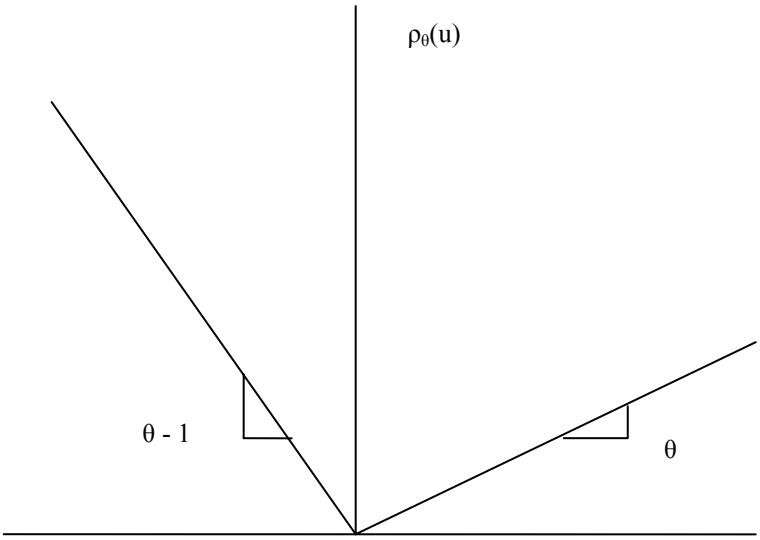
$$\min_{b\in\Re} n^{-1}\left\{\sum_{t\in\{t:y_t\geq x_t\beta\}}\theta\left|y_t - x_t\beta\right| + \sum_{t\in\{t:y_t<x_t\beta\}}(1-\theta)\left|y_t - x_t\beta\right|\right\} = \min \quad n^{-1}\sum_{i=1}^{n}\rho_\theta(y_t - x_t\beta)$$

(3.6)

em que ρ é a função "check" definida por

$$\rho_\theta(u) = \begin{cases}\theta\, u, & u \geq 0\\ (\theta - 1)u, & u < 0\end{cases} \qquad (3.7)$$

onde a função ρ_θ multiplica os resíduos por θ se eles forem não-negativos e por (θ – 1) caso contrário, para que desta forma sejam tratados assimetricamente. Graficamente, pode-se representar a função de perda, $\rho_\theta(u)$, da seguinte forma:



Pelos coeficientes estimados aos diferentes quantis, pode-se calcular a função quantil condicional, isto é, a distribuição empírica da variável dependente condicionada às covariáveis do modelo. Assim, no modelo linear com erros independentes e identicamente distribuídos (iid), a função quantil condicional é dada por:

$$Q_y(\theta \mid x) = x'\beta_\theta = x'\beta + Q_u(\theta) \qquad (3.8)$$

Para este caso, em que os erros são homocedásdicos, os coeficientes de cada quantil são simples deslocamentos paralelos uns aos outros, pois ambos possuem a mesma inclinação. O que irá diferi-los é o intercepto, dado por $\beta_0 + Q_u(\theta)$.

Na prática, os quantis de regressão são obtidos através da reformulação da função objetivo como um problema de programação linear através da introdução de 2n variáveis artificiais $\{u_i, v_i : 1, ...n\}$ para representarem as partes positivas e negativas do vetor de resíduos. Deste modo, tem-se:

$$\min_{(\beta, u, \nu) \in \Re \times \Re_{+}^{2n}} \left\{ \theta 1'_{n} u + (1-\theta) 1'_{n} v \middle| X\beta + u - v = y \right\} \qquad (3.9)$$

em que 1_n é um vetor de 1's.

Segundo Maciel (2001), citando Buchinsky[14], há importantes implicações quando a regressão quantílica é tratada como um problema de programação linear, pois garante que a estimativa de regressão quantílica seja obtida em um número finito de iterações simplex e permite robustez ao vetor de coeficientes estimado com relação às variáveis extremas (outliers).

3.4 As propriedades de Regressão Quantílica

Koenker & Bassett, em seu trabalho seminal em 1978, provaram a existência de importantes propriedades inerentes à técnica de estimação por meio de regressão quantílica, as quais referem-se às propriedades de invariância e robustez.

3.4.1 Equivariância

A propriedade de equivariância é importante no âmbito de estudos aplicados, pois esta permite que a escala da variável original possa ser alterada, sem que haja perda de coerência nas conclusões baseadas nos resultados estimados de regressão. Muitas situações são preferíveis a transformação dos dados (por exemplo, mudar a

[14] Buchinsky, M. (1997). Recent advances in quantile regression: A practical guideline for empirical research. Journal of Human Resources, 33 (1), 88-126.

escala da variável resposta de metros para quilômetros), pois desta forma facilita o tratamento computacional ou mesmo por gerar uma interpretação mais intuitiva[15].

Koenker & Bassett (1978) provaram que a técnica de estimação de regressão quantílica possui quatro importantes propriedades de equivariância, as quais serão reproduzidas a seguir. Para tal, $\hat{\beta}(\theta; y; X)$ indica o θ-ésimo quantil de regressão, $\theta \in (0,1)$; λ é um escalar maior que zero ($\lambda > 0$); $\gamma \in \Re^k$; e A é uma matriz k *x* k não-singular.

$$(1) \quad \hat{\beta}(\theta;\ \lambda y,\ X) = \lambda\, \hat{\beta}(\theta;\ y,\ X)$$

$$(2) \quad \hat{\beta}(1 - \theta;\ -\lambda y,\ X) = \lambda\, \hat{\beta}(\theta;\ y,\ X)$$

$$(3) \quad \hat{\beta}(\theta;\ y + X\gamma,\ X) = \hat{\beta}(\theta;\ y,\ X) + \gamma$$

$$(4) \quad \hat{\beta}(\theta;\ y,\ XA) = A^{-1}\, \hat{\beta}(\theta;\ y,\ X)$$

As propriedades (1) e (2) representam a equivariância em escala. Por exemplo, caso os dados de uma variável resposta sejam coletados em kilogramas (kg) e deseja-se transformá-los em gramas (g), basta multiplicar por 1000, gerando as seguintes estimativas: $\hat{\beta}(\theta; y[kg], X) = 1000 . \hat{\beta}(\theta; y[g], X)$. Propriedade (3) refere-se à equivariância de mudança ou de regressão. Propriedade (4) indica a equivariância em relação à matriz das variáveis explicativas.

[15] Koenker & Portnoy (1996) ilustraram a propriedade de equivariância com um exemplo de um modelo que analisa a temperatura de um líquido, y, em relação às variáveis explicativas, X. Uma vez que as medidas de temperatura de y estão mensuradas em grau Fahrenheit seria mais intuitivo mudar a escala para grau Celsius, pois é a medida de temperatura que estamos mais habituados. Assim, pela propriedade de equivariância, efetuando a mudança de grau Fahrenheit para grau Celsius, espera-se que

3.4.2 Invariância para transformações monotônicas

Além das propriedades de equivariância descritas na Seção 3.3.1,[16] a técnica de regressão quantílica (RQ) possui uma interessante vantagem em relação aos Mínimos Quadrados Ordinários (MQO), visto a primeira conter a interessante propriedade de *equivariância a transformações monotônicas*. Pela definição de função quantil para uma variável aleatória Y e assumindo que $h(.)$ seja uma função não-decrescente em $\Re$, tem-se a seguinte formalização:

$$Q_{h(Y)}(\theta) = h(Q_Y(\theta)) \tag{3.10}$$

Pela equação, os quantis da variável aleatória transformada h(y) são os quantis transformados da variável original Y. Um dos fatos que tornam a RQ atraente é exatamente esta propriedade, principalmente em modelos com censuras (*censoring models*)[17]. A esperança condicional, ao contrário, possui essa propriedade somente quando $h(.)$ é uma função linear ou em determinadas situações[18], sendo que para os demais casos essa não desfruta dessa propriedade, como mostra a inequação abaixo:

$$E\, h(y) \neq h(E(y)) \tag{3.11}$$

---

as estimativas de regressão mudem, pois mudou a escala, mas que sua interpretação permaneça invariante.

[16] As propriedades descritas nesta seção são compartilhadas tanto pelo método de regressão quantílica quanto por Mínimos Quadrados Ordinários.

[17] Ver Greene (2000) para uma introdução ao modelo com censura.

[18] Ver Koenker (2000) para um melhor detalhamento.

Para ilustrar a atratividade de RQ em relação a equivariância para transformações monotônicas, suponha um modelo linear com erros i.i.d., dado por:

$$y_i = x_i^T \beta + \varepsilon_i \tag{3.12}$$

em que $i \in \{1, \ldots, n\}$. Sendo a variável resposta não observada diretamente, mas utilizando-se um modelo com censura, pode-se observar $y_i^* = \max\{y_i, a\}$, em que $a \in \Re$ e representa o ponto censurado. Neste caso, métodos convencionais de regressão não geram estimativas consistentes, sendo que uma forma alternativa seria por meio do método de Máxima Verossimilhança (MV). Contudo, segundo Koenker & Portnoy (1996), reportando-se ao estudo de Goldberger (1983), existe a possibilidade de ocorrência de viés nas estimativas por MV caso a distribuição F não seja gaussiana. No entanto, como apontado por Powell (1986), quando se aplica a técnica de RQ, a função quantil condicional dependerá somente do ponto censurado e não da distribuição F, o que ocasiona à RQ a característica de ser uma técnica mais geral.

3.4.3 Robustez

A propriedade de robustez (ou sensibilidade) de um estimador vem sendo discutida por longo tempo. Uma questão que vem sendo debatida refere-se a qual estimador é mais robusto, o obtido pela média amostral ou pela mediana amostral?

As técnicas de regressão baseadas na média amostral tornaram-se populares desde que Gauss provou que, sobre presença de normalidade da distribuição de erros, a média amostral é o melhor estimador entre os estimadores não-viesados. Contudo,

caso ocorra uma única observação que seja distante o suficiente das demais observações, isso faz com que a média seja afetada significativamente, ocasionando viés à estimativa. A sensibilidade da média amostral a outliers enfatiza a fragilidade deste método em relação à robustez dos resultados obtidos por esta técnica. Por outro lado, o efeito de uma observação extrema (outlier) na mediana amostral é limitada, bem como aos demais quantis amostrais, porque são estimadas várias regressões para os diversos quantis amostrais, fazendo com que outliers afetem apenas localmente, isto é, a distância entre outliers e os quantis amostrais torna-se menor.

Uma forma comumente utilizada atualmente para medir a sensibilidade de um estimador, $\hat{\beta}$, refere-se à *função influência*. A partir desta, pode-se obter uma descrição da influência de uma contaminação da distribuição F em y (presença de outlier) sobre o estimador $\hat{\beta}$. Para o caso da média, a função influência indica que a contaminação da distribuição F em y é proporcional a y, implicando que, caso haja um ponto y afastado dos demais pontos, pode ocasionar uma média arbitrariamente distante dos valores iniciais de F. No caso da mediana, entretanto, a influência da contaminação em y é limitada pela constante $s(1/2) = 1/f\left(F^{-1}(1/2)\right)$ que é a função *sparsity* para a mediana. O resultado para a mediana pode ser estendido para os demais quantis, somente pela substituição da constante ½ pelo θ desejado, $s(\theta) = 1/f\left(F^{-1}(\theta)\right)$ (Koenker & Portnoy, 1996).

Assim, as estimativas geradas pelo método de regressão quantílica possuem a importante característica de serem robustas tanto para situações em que a função de distribuição de F em y for gaussiana (isto é, segue a distribuição normal), quanto para situações em que não se verifica a normalidade dos erros.

- CAPÍTULO IV -

Resultados

4.1 Modelo analítico: Efeito Quantílico do Tratamento

O reverencial analítico que será utilizado nesta dissertação refere-se ao emprego de uma variante da regressão quantílica, denominada de "Efeito Quantílico de um Tratamento" (quantile treatment effect). O estimador de efeitos quantílicos de um tratamento acomoda regressores exógenos, reduzindo a regressão quantílica a um caso especial quando o status de tratamento é exógeno. Assim, da mesma forma que os estimadores de variáveis instrumentais são equivalentes ao de mínimos quadrados ordinários (MQO) quando o status de tratamento é exógeno, o estimador do efeito quantílico de um tratamento (EQT) torna-se o estimador de regressão quantílica quando não há instrumentalização. Neste sentido, o EQT preserva todas as vantagens do modelo de regressão quantílica[19]. Desta forma, o uso de EQT permite analisar o efeito de um evento ou intervenção na distribuição de uma variável.

O Lema 2 (capítulo II) possibilita a generalização do efeito médio local de um tratamento (EMLT) para os quantis da variável resposta. Os quantis condicionais dos resultados potenciais para os *compliers* são dados por:

$$Q_\theta (Y_0 \mid X, D_1 > D_0) = X'\beta_\theta \,, \tag{4.1}$$

$$Q_\theta (Y_1 \mid X, D_1 > D_0) = \alpha_\theta + X'\beta_\theta, \tag{4.2}$$

[19] Estas formam descritas na seção 3.2.

em que θ é o índice do quantil pertencente a (0,1). A partir da variação exógena de D nos resultados potenciais (Y_0, Y_1) dados X e $D_1 > D_0$, obtém-se a seguinte função quantílica condicional para a população de *compliers*:

$$Q_\theta (Y \mid X, D, D_1 > D_0) = \alpha_\theta D + X'\beta_\theta \quad (4.3)$$

onde y_i é a variável resposta e X_i é uma matriz de variáveis explicativas. O coeficiente α_θ tem uma interpretação de causalidade porque é a diferença entre o θ-ésimo quantil de Y_1 e Y_0 para os *compliers*. Deste modo, α_θ indica se houve mudança no rendimento familiar por causa da presença do terceiro filho. A distribuição marginal representa, então, o centro da análise visto ser somente esta que pode ser identificada pela aleatoriedade da população de *compliers*[20].

Os parâmetros da função quantílica condicional podem ser expressos da seguinte forma:

$$(\alpha_\theta, \beta_\theta) \equiv \text{argmin}_{(\alpha,\beta)} E [\rho_\theta (Y - \alpha D - X'\beta) \mid D_1 > D_0], \quad (4.4)$$

onde $\rho_\theta(\lambda)$ é a função "check", dado por $\rho_\theta(\lambda) = [\theta - I(\lambda < 0)]\lambda$ para qualquer λ real. A formulação em (4.4) não pode ser utilizada diretamente, uma vez que a população de *compliers* não é identificável. Assim, asseguradas pelo Lema 2 do Capítulo II, as estimativas de α_θ e β_θ por EQT podem ser obtidas através da minimização do seguinte problema de regressão quantílica ponderada:

[20] Fazendo uso do exemplo descrito no capítulo II, significa dizer que podemos saber se o programa de treinamento deslocou para cima o percentil 0,10 da distribuição de renda, mas não possibilita saber se as pessoas que originalmente estavam na percentil 0,10 da distribuição de renda tiveram um aumento em seus rendimentos.

$$(\alpha_\theta, \beta_\theta) = \text{argmin}_{\{\alpha, \beta\}} \; E\,[k \,.\, \rho_\theta\,(Y - \alpha D - X\beta)], \tag{4.5}$$

em que *k* é a função que identifica os *compliers*. Desta forma, (4.5) é o estimador natural de α_θ e β_θ, pois neste caso tem-se a função objetivo globalmente convexa em α_θ e β_θ garantida pela minimização dos erros da função *check*. Entretanto, *k* é somente positivo para os casos em que Z = D; caso contrário, este assume valores negativos, ocasionado uma função não-convexa. Deste modo, é necessário o uso de um algoritmo que minimize este problema. A formulação que segue utiliza o processo de iteração sobre (Y, D, X) para tornar a formulação em (4.5) computável, isto é, que tenha uma representação em programação linear.

$$(\alpha_\theta, \beta_\theta) = \text{argmin}_{\{\alpha, \beta\}} \; E\,[k_v \,.\, \rho_\theta\,(Y - \alpha D - X\beta)] \tag{4.6}$$

onde

$$k_v = 1 - \frac{D.(1 - \nu_0(U))}{1 - \pi_0(X)} - \frac{(1 - D).\nu_0(U)}{\pi_0(X)}, \tag{4.7}$$

para $\pi_0 = E[Z \mid X]$; $\nu_0(U) = E[Z \mid U] = P(Z = 1 \mid Y, D, X)$; e $U = (Y, D, X)$. Aplicando este algoritmo, $k_v(U)$ pode ser interpretado como a probabilidade de ocorrência de *compliers* dada a variável resposta, o tratamento e as variáveis explicativas do modelo, isto é, $k_v(U) = P(D_1 > D_0)$. Assim, uma vez estimado $k_v(U)$, este não pode conter valores negativos, pois trata-se de uma probabilidade condicional. Um mínimo

global, então, pode ser obtido pela função *check* em um número finito de iterações simplex.

Segundo Abadie, Angrist & Imbens (2002), não há perda em eficiência assintótica pelo uso da estimativa de $k_v(U)$, ao invés do proposto em (4.5). Ambas as estratégias produzem estimadores com a mesma distribuição assintótica, com o diferencial que em (4.6) o manuseio computacional é mais atraente, pois este é operacionalizado no âmbito de um problema convexo.

O primeiro passo para estimar $k_v(U)$ é proceder com a estimação de $\pi_0(X)$ e $v_0(X)$. Definindo $\pi_0(X) = E[Z = 1 \mid X]$ e se $Z \perp\!\!\!\perp X$, tem-se que $E[Z = 1 \mid X] = E[Z]$. A estimação de $v_0(U) = E[Z \mid Y, D, X]$ é obtida pelo método não-paramétrico de *power series*[21], a qual é dada pela projeção em MQO de Z sobre as iterações de Y, D, X.

4.2 Dados

Para implementar a estratégia de estimação a qual a presente dissertação se propôs, são utilizadas informações da Pesquisa Nacional de Amostragem Domiciliar (PNAD), coletada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), para o ano de 1999.

A amostra selecionada consiste nas mulheres de faixa etária entre 21 e 35 anos, as quais satisfazem os seguintes requisitos: i) tenham pelo menos dois filhos; ii) que a renda familiar (descontada a renda dos filhos) seja maior do que zero; iii) que o primogênito tenha idade inferior a 18 anos[22]; iv) que todas as informações de interesse

[21] Este método é utilizado em Abadie, Angrist & Imbens (2002), Abadie (2001), entre outros.
[22] Idade bastante propensa a constituir ou mudar-se para uma nova família (ver Angrist e Evans, 1998).

sejam completadas no banco de dados da PNAD. Nesta perspectiva, a amostra se restringiu a aproximadamente quatorze mil famílias.

Nesta aplicação, Y é o log da renda familiar[23] (excluindo-se a renda dos filhos) para a amostra de mulheres com dois ou mais filhos, D indica mulheres com três ou mais filhos (> dois filhos) e Z é o instrumento[24], o qual indica se os dois primeiros filhos são ambos meninos ou ambas meninas. A matriz de covariáveis consiste em uma constante, idade da mãe, idade da mãe quando esta teve o primeiro filho, uma *dummy* referente a mãe ser chefe de família, *dummies* educacionais[25], uma *dummy* racial (branco=1), *dummies* regionais, *dummy* para áreas metropolitanas, *dummy* para área urbana e uma *dummy* para o caso do primeiro filho ser do sexo masculino.

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 conseqüência de haver uma 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 (este 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 ter viés de seleção em conseqüência da idade máxima do primogênito ter sido limitada a ser menor de 18 anos. Para o grupo selecionado (mulheres com dois ou mais filhos na faixa etária de 21 a 35 anos), o(a) filho(a) mais velho(a) de aproximadamente 86%

[23] A renda familiar usada exclui os agregados.

[24] Nesta dissertação também utilizou-se o instrumento "gêmeos", o qual indicava a ocorrência do nascimento de gêmeos na primeira gravidez da mulher, como sugerido pelo artigo de Angrist e Evans (1998). Contudo, para o caso brasileiro, a amostra das famílias que tiveram gêmeos na primeira gravidez e que preencheram os demais quesitos expostos nesta seção é muito pequena, o que impossibilitou estimar com precisão o efeito da criação de um terceiro filho na distribuição de renda das famílias.

[25] Optou-se por construir as *dummies* educacionais a partir de intervalos de anos de estudos, os quais representam uma proxy do grau de instrução do indivíduo. Por exemplo, o intervalo 1 a 3 anos de estudo indica que o indivíduo possui menos que o primário. Os demais intervalos são definidos de forma similar.

das mães tinha idade inferior a 18 anos. Embora estas mães possam parecer pertencerem a um grupo jovem de alta fecundidade e não usual, convém salientar que 39,64% de todas as mulheres na faixa etária 21-35 anos estão neste grupo. O percentual é similar para as mulheres nesta faixa etária com mais de dois filhos (39,95%).

Tabela 1: Medidas de Fecundidade e Oferta de Trabalho das Mães

- **Mulheres com idade entre 21 e 35 anos**

  Média de filhos nascidos 1,44
  Percentagem com dois filhos ou mais 39,64%
  Percentagem que trabalhou na semana de referência*..... 53,69%
  Total de observações 44.462

- **Mulheres com idade entre 36 e 50 anos**

  Média de filhos nascidos 2,79
  Percentagem com dois filhos ou mais 59,57%
  Percentagem que trabalhou na semana de referência*..... 57,12%
  Total de observações 33.513

- **Mulheres com idade entre 21 e 35 anos com dois ou mais filhos, sendo o primogênito de idade inferior a 18 anos**

  Média de filhos nascidos 2.76
  Percentagem com mais de dois filhos 39,95%
  Percentagem que trabalhou na semana de referência*..... 45,73%
  Total de observações 13.988

* De 19 a 25 / 09 / 99.

A tabela 1 mostra as estatísticas descritivas referentes a taxa de fecundidade e a participação na força de trabalho das mulheres com idade entre 21-35 e 36-50 anos. Para o primeiro grupo, a média de filhos é de 1,44. Quando condicionamos este grupo de mães as condições de ter dois ou mais filhos e que o(a) filho(a) mais velho(a) tenha idade inferior a 18 anos, as quais correspondem a 31,46% de todas as mulheres da

amostra com a respectiva idade (21-35 anos), esta média sobe para 2,76. No segundo grupo (36-50), a média de filhos é de 2,79. Vê-se, também, que 53,69% das mulheres na faixa etária 21-35 anos estão na força de trabalho, enquanto que este percentual é de 57,12% para mulheres em idade no intervalo 36-50.

A tabela 2 exibe as estatísticas descritivas das covariáveis, instrumento e variável resposta. A covariável 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". Dentre as mulheres que tiveram dois filhos, 40% tiveram um terceiro filho e aproximadamente 50% tiveram os dois primeiros filhos do mesmo sexo, enquanto que apenas 1% tiveram gêmeos na primeira gravidez[26]. Observa-se pela tabela 2 que 51% dos primeiros nascimentos destas famílias foram de crianças do sexo masculino. Em 14% dos casos, as mulheres são as pessoas de referência da família, isto é, são as responsáveis pelas famílias. A renda média destas mulheres é de R$ 137,09, sendo que a renda familiar, descontada a renda dos filhos que trabalhavam no período de referência da coleta dos dados, sobe para R$ 668,29. Em 79,96% dos casos, as mulheres residem em áreas urbanas e a grande maioria (64%) possui escolaridade inferior ao primeiro grau completo[27]. Pouco mais de 3% obtiveram o terceiro grau completo (graduação).

[26] 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. Vê-se que para o caso brasileiro estas famílias são pouco representativas, o que ocasionou inviabilidade deste instrumento para o propósito da aplicação desta dissertação.

[27] Cada série concluída com aprovação corresponde a um ano de estudo, ou seja, a repetência está sendo controlada.

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
Filhos nascidos	2.762537	1,198837
Mais de dois filhos	0.3995568	-
Mesmo sexo	0.4976408	-
Gêmeos	0.0085788	-
Chefe de família	0.140835	-
Primeiro filho homem	0.5120103	-
Idade	29.57	3,932764
Idade na primeira gravidez	20.27	3,334896
Renda da mãe	137.09	380,89
Renda familiar	668.29	1028,53
Raça (=1 se branca)	0.4838433	-
Urbana	0.7996854	-
Rural	0.2003146	-
Metropolitana	0.3655991	-
NE	0.3640263	-
NO	0.0996568	-
SE	0.1940234	-
SUL	0.186517	-
CO	0.1557764	-
Nenhuma instrução	0.0896483	-
De 1 a 3 anos de estudo	0.1592079	-
De 4 a 7 anos de estudo	0.3930512	-
De 8 a 10 anos de estudo	0.1644981	-
De 11 a 14 anos de estudo	0.1563483	-
15 ou mais anos de estudo	0.0372462	-

4.3 Aplicação

Se a fecundidade e os rendimentos são determinados conjuntamente, como sugerido pela teoria econômica (ver Browning, 1992) faz com que estimativas desta relação pelo método de MQO ou de regressão quantílica não tenham uma interpretação causal. A estimação por meio de variáveis instrumentais do efeito quantílico de um tratamento contorna este problema.

O primeiro passo para proceder com a estimação do efeito quantílico do tratamento é estimar o peso k, que, em outros termos, significa encontrar a probabilidade de ocorrência de *compliers*. Entretanto, para estimar k, deve-se primeiro estimar π_0 (X) = E[Z | X] e ν_0(U) = E[Z | Y, D, X]. Empiricamente, constatou-se que Z não é independente de todo conjunto de variáveis explicativas, sendo estatisticamente significativa a relação entre o instrumento (Z) e a *dummy* da ocorrência do primeiro filho ser do sexo masculino (Dpfh). Deste modo, π_0 (X) = E[Z | Dpfh]. O ν_0(U) é estimado não-parametricamente pelo método de *power series*[28]. Assim, pelo fato de Z ser fortemente correlacionado com o tratamento (D) faz com que D seja iteragido em todos os termos da série. Do mesmo modo, como a *dummy primeiro filho homem* é correlacionada com Z, esta também é iteragida na série. Assim, ν_0(U) é estimado pela projeção em MQO de Z nas iterações de Y, D, X. Desta forma, estimam-se o seguinte modelo de regressão:

$$\text{Instrumento} = \text{constante} + \sum_{i=1}^{n} (\textit{iteração})^{i} \qquad i = 1, 2, ...,10$$

em que a iteração, para o caso em que D = 1 e Dpfh = 1, é definida por: Y^1* D = 1 * Dpfh = 1, quando i =1; Y^1* D = 1 * Dpfh = 1 + Y^2* D = 1 * Dpfh = 1, quando i = 2 e assim por diante, onde D é o tratamento,Y é o log da renda familiar e Dpfh é a *dummy* para o caso do primeiro filho ser do sexo masculino. Os termos da iteração também irão constar na constante. No caso desta aplicação, teremos quatro modelos de iteração para proceder com a estimativa de ν_0(U), os quais são: quando D = 1 e Dpfh=1 (caso exposto acima); D = 1 e Dpfh = 0 ; D = 0 e Dpfh = 1; D = 0 e Dpfh = 0.

[28] Para maiores detalhes ver Abadie (2002).

Para obter a melhor iteração, utilizam-se os critérios Akaike Information criterion (AIC) e Bayesian Information Criterion (BIC) e opta-se pela iteração que obtiver melhor AIC e BIC. Foram estimadas 10 iterações, sendo a primeira a com menor AIC e BIC, como mostra a Tabela 3.

$$AIC = \log\sigma^2 + \frac{p.\log n}{n} \quad ,$$

$$BIC = \log\sigma^2 + \frac{n+p}{n-p+2}$$

Tabela 3: Resultados do AIC e BIC para as iterações de $\nu_0(U)$

Iterações	D=1, Dpfh=1		D=1, Dpfh=0		D=0, Dpfh=1		D=0, Dpfh=0	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
1°	**-0.8178**	**0.18096**	**-0.81277**	**0.18600**	**-0.8608**	**0.1378**	**-0.8283**	**0.17047**
2°	-0.8170	0.18116	-0.81202	0.18621	-0.8601	0.1381	-0.8277	0.17051
3°	-0.8163	0.18136	-0.81142	0.18628	-0.8593	0.1383	-0.8270	0.17061
4°	-0.8155	0.18157	-0.81066	0.18649	-0.8586	0,1384	-0.8263	0.17081
5°	-0.8149	0.18169	-0.80999	0.18663	-0.8579	0,1386	-0.8255	0.17102
6°	-0.8142	0.18180	-0.80924	0.18684	-0.8572	0,1388	-0.8248	0.17123
7°	-0.8142	0.18189	-0.80851	0.18703	-0.8572	0,1391	-0.8240	0.17144
8°	-0.8135	0.18198	-0.80850	0.18704	-0.8558	0,1391	-0.8233	0.17165
9°	-0.8128	0.18212	-0.80800	0.18704	-0.8558	0,1392	-0.8233	0.17164
10°	-0.8128	0.18213	-0.80776	0.18724	-0.8558	0,1395	-0.8226	0.17185

Estimado π_0 (X) e $\nu_0(U)$ utiliza-se o algoritmo proposto em (4.7) para identificar a população de *compliers*. Com k estimado, pode-se proceder com EQT.

A investigação começa por mostrar (Tabela 4) os resultados obtidos pelas estimativas de MQO e MQ2E, com o intuito de comparação. A estimativa por MQO para o efeito de ter o terceiro filho é de uma redução de aproximadamente 12% na renda familiar. Pelo método de MQ2E foi encontrado um efeito negativo de 46% referente ao terceiro filho na renda familiar, com nível de significância de 6,2%.

Estimativas de regressão quantílica (RQ) são reportadas na Tabela 5. Estas mostram um efeito para o caso da mediana de -12,64%, sendo o menor efeito encontrado no quantil 0,75 (-8,97%) e os maiores efeitos nos quantis inferiores (θ =

0,10 e θ = 0,25) e no quantil 0,90. O maior impacto (em termos absolutos) encontra-se no quantil 0,10 com uma redução nos rendimentos da família de 16,68%.

A Tabela 6 mostra os resultados extraídos das estimativas por meio do EQT. A estimativa de EQT do efeito de mais de dois filhos para a mediana foi de -13,56%. Esta é maior (em valor absoluto) do que a encontrada na estimativa de regressão quantílica simples (Tabela 5). Os resultados de EQT possuem um padrão semelhante ao apontado pela Tabela 5 (RQ) no que concerne ao fato das famílias pertencentes ao quantil 0,10 (representam as famílias de baixa) serem as que incorrem em uma maior redução na renda familiar. Quando no movemos para a cauda direita da distribuição condicional da renda familiar, em ambos métodos, as estimativas decrescem (em valor absoluto), chegando a seu valor mínimo pelo método de Regressão Quantílica no quantil 0,75 (-8,97%) e por EQT ao quantil 0,50 (-13,56%).

Um fato que deve ser salientado diz respeito às estimativas por EQT serem maiores (em valor absoluto) do que as mesmas obtidas por RQ. De fato, o viés gerado pela endogeneidade na relação das variáveis em questão faz com que a estimativa por RQ subestime o efeito causal do terceiro filho na distribuição condicional da renda familiar. Assim, além do efeito causal do terceiro filho, os resultados de RQ possivelmente estão refletindo as grandes desigualdades de renda no Brasil e significativas diferenças na taxa de fecundidade entre famílias pobres e ricas.

Pela amostra desta dissertação, 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 este 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. Isto demonstra que quase a

metade das famílias que são mais numerosas (três ou mais filhos), encontram-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 afetar a oferta de trabalho das mesmas e, conseqüentemente, a renda familiar. Por sua vez, o resultado para as famílias mais abastadas podem ser uma conseqüência de as mesmas optarem por investir mais na qualidade da criação dos filhos, com resultado de uma maior dedicação de seu tempo aos mesmos.

Com respeito ao retorno a educação há algumas interessantes constatações que podemos inferir a partir dos resultados gerados. De modo geral, o retorno a educação é maior para os indivíduos que possuem maior grau de escolaridade. Por exemplo, o retorno a educação é de 0,10 para os indivíduos que possuem menos que o primário, elevando-se para 1,728 para os indivíduos com 15 anos ou mais de estudo (todos analisados na mediana amostral), o que demonstra a importância da educação na determinação dos rendimentos.

Duas outras características podem ser extraídas dos resultados: i) o incremento na renda familiar provido pela educação para os indivíduos com alto nível educacional cresce continuamente da cauda esquerda para a cauda direita da distribuição condicional da renda familiar, enquanto que para indivíduos com nível educacional baixo este padrão não se verifica. Por exemplo, quando condicionamos os indivíduos com 15 ou mais anos de estudo, o retorno a educação é 1,17 para os indivíduos pertencentes ao quantil 0,10 e eleva-se continuamente ao longo dos quantis, passando a atingir 2,048 no quantil 0,9; no caso de nível educacional mais baixo (menos que o primário), por sua vez, temos que um indivíduo que se encontra na quantil 0,10 da distribuição tem retorno de 0,104, ao passo que se estivesse no

quantil 0,25 teria um retorno de 0,059 ou, ainda, para o caso da mediana, o retorno estimado seria de 0,101; ii) existe maior variabilidade do retorno a educação para indivíduos que possuem um nível mais elevado de educação, sendo menor esta variabilidade quando se analisa indivíduos com nível educacional mais baixo. Por exemplo, para pessoas com 15 anos de estudo ou mais a diferença entre os salários mais baixos (quantil 0,10) e mais altos (quantil 0,90) é bastante significativa, a qual representa 0,877. No entanto, condicionando aos indivíduos que possuem instrução inferior ao primário, esta diminui para 0,06.

Referente as demais variáveis do modelo observamos que os sinais se comportam como o esperado: i) há um incremento na renda quando a mulher é branca, sendo este crescente (exceto pelo quantil 0,25, o qual apresenta uma leve redução relativamente ao quantil 0,1) a medida que nos deslocamos da cauda esquerda para a cauda direita da distribuição condicional da renda familiar; ii) o efeito da variável binária relacionada a área urbana na renda familiar é positivo, variando entre 0,25 (quantil 0,10) e 0,31 (quantil 0,90) ; iii) morar na região metropolitana também contribui para acrescer a renda familiar; iv) habitar na região Nordeste (dummy excluída) implica em ter menores rendimentos que nas demais regiões (mantidas as demais características constantes); v) a idade da mãe se relaciona positivamente com a renda familiar, indicando que mulheres com mais idade tendem a aumentar a renda familiar. No entanto, vi) a idade da mãe quando esta teve o primeiro filho se relaciona inversamente com a renda familiar, ou seja, quanto mais nova for a mãe na primeira gravidez[29], maior (em termos absolutos) será o impacto negativo na renda familiar. Este resultado é de aproximadamente -1%, sendo praticamente igual em todos os quantis da distribuição da variável resposta; vii) referente a *dummy* chefe de família,

[29] A qual resultou na geração do primeiro filho.

esta sinalizou que mulheres que são responsáveis pela família tem sua renda familiar reduzida em pouco mais de 50%. Este alto coeficiente (em termos absolutos) provavelmente está indicando a escassez de tempo da mulher chefe de família em se aperfeiçoar (aumentar seu capital humano), reduzindo suas chances de auferir uma remuneração melhor.

- CAPÍTULO V -

Conclusão

Esta análise reporta estimativas do efeito do terceiro filho nos quantis condicionais do log da renda familiar. Foi aplicado um novo estimador para o efeito de um tratamento endógeno na distribuição condicional da variável resposta. O estimador do efeito quantílico do tratamento (EQT) pode ser usado para determinar como uma intervenção afeta a distribuição de uma variável para indivíduos em que o status de tratamento é influenciado por um instrumento binário. A partir deste, pode-se estimar uma bem definida aproximação do efeito causal de interesse. O estimador de EQT do efeito de mais de dois filhos na distribuição condicional do log da renda familiar indica interessantes e importantes diferenças aos diferentes quantis condicionais. Assim, podem-se sumarizar as seguintes conclusões:

- Estimar a relação entre o número de filhos e a renda familiar por MQO deve ser visto com cautela, pois tal método não é apropriado para casos em que há endogeneidade na relação entre as variáveis de interesse. Mesmo o coeficiente do terceiro filho em MQO ser estatisticamente significativo, este não estima precisamente o efeito causal, pois é bastante provável que este contenha viés.

- Ao contrário de MQO, o método de MQ2E lida como o problema da endogeneidade, pois se utiliza de uma variável instrumental que é correlacionada com o tratamento (variável explicativa) e não-correlacionada

com a variável resposta, contornando, deste modo, o problema da endogeneidade nesta relação. A ressalva que se faz refere-se ao fato que as estimativas obtidas por MQ2E do efeito do tratamento (terceiro filho) na renda familiar são estimadas para a média amostral, o que para o caso desta aplicação gera uma visão incompleta da distribuição dos efeitos, pois o impacto do terceiro filho provavelmente não é homogêneo para a distribuição condicional da renda familiar.

- Uma importante característica evidenciada por este trabalho refere-se à assimetria na resposta do impacto do terceiro filho na distribuição condicional da renda familiar. Enquanto no quantil 0,10 a presença do terceiro filho reduz a renda familiar em 20%, este impacto decresce (em valor absoluto) para aproximadamente -13,5% para o quantil 0,50 (ver tabela 6). Padrão similar se verifica para as estimativas geradas por regressão quantílica sem ponderação (a qual utilizou o mesmo peso para a população amostral: *compliers, always-takers* e *never-takers*); contudo, esta subestima este impacto, provavelmente por causa da endogeneidade entre a variável resposta e o tratamento.

- As estimativas produzidas por EQT mostraram que o terceiro filho provoca redução na renda familiar para todos os quantis condicionais. Deste modo, estes resultados podem[30], de certa forma, corroborar com a Teoria da Família, a qual sustenta a hipótese de que cada vez mais as famílias optam por famílias pequenas, visto o elevado custo de oportunidade de ter filhos, bem como a difícil tarefa de conciliar família grande e oferta de trabalho. Assim, a escolha

[30] Juntamente com a queda da taxa de natalidade verificada pelo IBGE.

entre aumentar o número de filhos e intensificar o investimento em capital humano do(s) filho(s) tende à última opção, na argumentação da Teoria da Família.

- O EQT é maior para os quantis inferiores da distribuição do log da renda familiar, indicando que as famílias com menor poder aquisitivo (quantis 0,10 e 0,25) têm sua renda familiar decrescida em maior proporção, aproximadamente 20% e 18%, respectivamente, quando estas têm o terceiro filho.

- Ainda acerca das implicações da Teoria da Família, cabe fazer algumas distinções que nos parecem relevantes. Os resultado de EQT indicam que as famílias de mais baixa renda (quantis 0,10 e 0,25) são as que têm sua renda familiar mais reduzida quando o terceiro filho é gerado. A intuição nos leva a acreditar que a causa do efeito ser de maior magnitude para as famílias mais pobres seja pela necessidade de cuidar dos filhos, dedicando assim menos tempo ao trabalho ou mesmo permanecendo fora do mercado de trabalho. Para as famílias mais ricas, a quais também têm sua renda familiar reduzida (em torno de 15%), por sua vez, pode indicar uma escolha das mães nestas famílias de dedicarem maior parte de seu tempo à criação dos filhos, para assim elevarem o capital investido nos mesmos.

- A evidência empírica do EQT para o caso brasileiro possui um padrão peculiar se comparado com as estimativas geradas com dados estadunidenses. Abadie, Angrist & Imbens (1998) estimaram o EQT e verificaram que as famílias

pertencentes aos quantis mais baixos são as que têm a renda familiar mais reduzida, sendo o quantil 0,90 muito pouco afetado, se comparado relativamente aos demais quantis. No caso brasileiro, o impacto do terceiro filho na renda familiar decresce (em valor absoluto) à medida que o quantil é aumentado. Entretanto, no quantil 0,75 o padrão é alterado, pois ao invés de decrescer, como era de se esperar pelo padrão nos demais quantis, este aumenta. Este possivelmente reflete o alto custo de oportunidade com que os pais (principalmente a mãe, como sugerido pela Teoria da Família) se deparam ao optar em aumentar o tamanho da família[31].

- Outro aspecto importante apontado pelos dados brasileiros referem-se a magnitude da queda na renda resultante de um terceiro filho, a qual é maior no caso brasileiro que nos EUA[32].

- Em geral, todos os métodos econométricos utilizados nesta dissertação (MQO, MQ2E, RQ e EQT) indicam que a geração e criação do terceiro filho provoca uma redução na renda familiar, independente do nível de renda da família.

[31] Esta diferença ao que concerne as estimativas obtidas a partir de dados estadunidenses e brasileiros para o quantis superiores (0,75 e 0,90), talvez resulte de fatores culturais, de estilo de vida e estruturais entre os dois países. Pode acontecer de as 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 EUA que permitam maior eficiência na criação dos filhos, permitindo que as mulheres tenham mais tempo para dedicarem-se ao trabalho além das horas alocadas para a família.
[32] Os resultados para os EUA são encontrados no trabalho de Abadie et al. (1998).

Tabela 4. Resultados por MQO e MQ2E. Nota: coeficientes em negrito e P – valores na coluna seguinte

Log renda família	MQO		MQ2E	
Constante	4.319989	0,000	4.513871	0,000
> 2 filhos	-.1239877	0,000	-.4606341	0.062
Idade	.0348818	0,000	.0463024	0,000
1° filho homem	.1246003	0,000	.1322716	0,000
Chefe de família	-.5428789	0,000	-.550131	0,000
Raça	.1897056	0,000	.1699942	0,000
Urbana	.2843939	0,000	.2613865	0,000
Metropolitana	.1177788	0,000	.1113588	0,000
NO	.2361176	0,000	.2578581	0,000
SE	.3993045	0,000	.3886893	0,000
SUL	.267278	0,000	.2576491	0,000
CO	.3364519	0,000	.3110856	0000
Idade 1° filho	-.0127703	0,000	-.0276044	0,013
1 a 3 anos de estudo	.0929708	0,000	.0749263	0,012
4 a 7 anos de estudo	.2912688	0,000	.2408172	0,000
8 a 10 anos de estudo	.6076141	0,000	.52955	0,000
11 a 14 anos de estudo	1.06116	0,000	.960147	0,000
15 ou mais anos de estudo	1.56744	0,000	1.470148	0,000

Tabela 5. Resultados por Regressão Quantílica. Nota: coeficientes em negrito
e P – valores em parênteses

log renda fam.	θ = 0.1		θ = 0.25		θ = 0.5		θ = 0.75		θ = 0.9	
Constante	**3.857124**	(0.000)	**4.120429**	(0.000)	**4.384278**	(0.000)	**4.693799**	(0.000)	**4.887028**	(0.000)
> 2 filhos	**-.166863**	(0.000)	**-.157965**	(0.000)	**-.126482**	(0.000)	**-.089756**	(0.000)	**-.132936**	(0.000)
Idade	**.0274545**	(0.000)	**.0291257**	(0.000)	**.0347981**	(0.000)	**.0373002**	(0.000)	**.0445781**	(0.000)
1° filho homem	**.2542259**	(0.000)	**.1599753**	(0.000)	**.0911127**	(0.000)	**.04131**	(0.007)	**.067636**	(0.000)
Chefe de família	**-.4763065**	(0.000)	**-.521489**	(0.000)	**-.534699**	(0.000)	**-.526182**	(0.000)	**-.530534**	(0.000)
Raça	**.1124132**	(0.000)	**.151767**	(0.000)	**.1558719**	(0.000)	**.2106338**	(0.000)	**.2437808**	(0.000)
NO	**.2099861**	(0.000)	**.2173829**	(0.000)	**.195727**	(0.000)	**.2607614**	(0.000)	**.311798**	(0.000)
SE	**.4319044**	(0.000)	**.4331138**	(0.000)	**.4464827**	(0.000)	**.3934851**	(0.000)	**.3161859**	(0.000)
SUL	**.290345**	(0.000)	**.3218049**	(0.000)	**.3063295**	(0.000)	**.2437295**	(0.000)	**.1977465**	(0.000)
CO	**.3286948**	(0.000)	**.3331869**	(0.000)	**.3165092**	(0.000)	**.3314654**	(0.000)	**.3282061**	(0.000)
Urbana	**.2378753**	(0.000)	**.2573626**	(0.000)	**.2927438**	(0.000)	**.2736812**	(0.000)	**.276727**	(0.000)
Metropolitana	**.0980321**	(0.000)	**.1113615**	(0.000)	**.0991736**	(0.000)	**.1395582**	(0.000)	**.1466873**	(0.000)
Idade 1° filho	**-.0132397**	(0.000)	**-.011642**	(0.000)	**-.015310**	(0.000)	**-.014278**	(0.000)	**-.016583**	(0.000)
1 a 3 anos de estudo	**.0692999**	(0.069)	**.0765367**	(0.020)	**.0982508**	(0.001)	**.0775618**	(0.018)	**.1267507**	(0.001)
4 a 7 anos de estudo	**.2285269**	(0.000)	**.2324977**	(0.000)	**.2788458**	(0.000)	**.2849224**	(0.000)	**.3482753**	(0.000)
8 a 10 anos de estudo	**.4636409**	(0.000)	**.491161**	(0.000)	**.5902899**	(0.000)	**.6435026**	(0.000)	**.7491805**	(0.000)
11 a 14 anos de estudo	**.7984079**	(0.000)	**.8752345**	(0.000)	**1.062595**	(0.000)	**1.198389**	(0.000)	**1.298232**	(0.000)
15 ou mais anos de estudo	**1.069488**	(0.000)	**1.29857**	(0.000)	**1.611991**	(0.000)	**1.860037**	(0.000)	**1.97405**	(0.000)

Tabela 6. Resultados por Efeito Quantílico do Tratamento (EQT). Nota: coeficientes em negrito e P – valores na coluna seguinte

log renda fam.	θ = 0.1		θ = 0.25		θ = 0.5		θ = 0.75		θ = 0.9	
Constante	**3.82211**	0.000	**4.10873**	0.000	**4.41270**	0.000	**4.62333**	0.000	**5.02012**	0.000
> 2 filhos	**-0.20228**	0.000	**-0.18210**	0.000	**-0.13569**	0.000	**-0.15530**	0.000	**-0.14851**	0.000
Idade	**0.02940**	0.000	**0.03295**	0.000	**0.03556**	0.000	**0.04225**	0.000	**0.04066**	0.000
1° filho homem	**0.39135**	0.000	**0.31888**	0.000	**0.25998**	0.000	**0.23925**	0.000	**0.26762**	0.016
Chefe de família	**-0.50879**	0.000	**-0.54629**	0.000	**-0.53599**	0.000	**-0.52556**	0.000	**-0.52905**	0.000
Raça	**0.17026**	0.000	**0.16330**	0.000	**0.19781**	0.000	**0.24627**	0.000	**0.26257**	0.000
NO	**0.22817**	0.000	**0.19294**	0.000	**0.20256**	0.000	**0.28362**	0.000	**0.30545**	0.000
SE	**0.45664**	0.000	**0.42028**	0.000	**0.38410**	0.000	**0.30517**	0.000	**0.21604**	0.000
SUL	**0.31086**	0.000	**0.29610**	0.000	**0.24357**	0.000	**0.14313**	0.000	**0.09910**	0.019
CO	**0.37256**	0.000	**0.32691**	0.000	**0.31089**	0.000	**0.33395**	0.000	**0.27130**	0.000
Urbana	**0.25700**	0.000	**0.28885**	0.000	**0.30698**	0.000	**0.29779**	0.000	**0.31308**	0.000
Metropolitana	**0.11064**	0.000	**0.11918**	0.000	**0.10027**	0.000	**0.14053**	0.000	**0.11348**	0.000
Idade 1° filho	**-0.01222**	0.0002	**-0.01187**	0.0001	**-0.01386**	0.000	**-0.01263**	0.000	**-0.01429**	0.001
1 a 3 anos de estudo	**0.10403**	0.0173	**0.05903**	0.0717	**0.10171**	0.001	**0.10441**	0.003	**0.16794**	0.000
4 a 7 anos de estudo	**0.28304**	0.000	**0.24781**	0.000	**0.29954**	0.000	**0.31461**	0.000	**0.41627**	0.000
8 a 10 anos de estudo	**0.51110**	0.000	**0.54716**	0.000	**0.62644**	0.000	**0.68469**	0.000	**0.85796**	0.000
11 a 14 anos de estudo	**0.85392**	0.000	**0.93284**	0.000	**1.09286**	0.000	**1.20940**	0.000	**1.39946**	0.000
15 ou mais anos de estudo	**1.17182**	0.000	**1.39013**	0.000	**1.72884**	0.000	**1.87211**	0.000	**2.04808**	0.000

BIBLIOGRAFIA

1) ABADIE, A. (1997). Identification of treatment effects in models with covariates. MIT Department of Economics, mimeo.

2) ABADIE, A. (2001). Semiparametric instrumental variable estimation of treatment response models. http://ksghome.harvard.edu/~.aabadie. academic.ksg/

3) ABADIE, A., ANGRIST, J.D. & IMBENS, G.W (1998). Instrumental variables estimation of quantile treatment effects. NBER Working Paper, 229, 1-28.

4) ABADIE, A., ANGRIST, J.D. & IMBENS, G.W (2002). Instrumental variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. Econometrica 70, 91-117

5) ANGRIST, J.D. & EVANS, W.N. (1998). Children and their parent's labor supply: Evidence from exogenous variation in family size. American Economic Review, 88(3), 450-477.

6) BECKER, Gary S. (1960). An economic analysis of fertility. In: Demographic and Economic Change in Developed Countries. Universities-National Bureau Conference Series 11. Princeton.

7) BECKER, Gary S. & LEWIS, Gregg H. (1973). On the interaction between the quantity and quality of children. Journal of Political Economy. V. 81, N. 2, S279-S288.

8) BEN-PORATH, Yovan (1973). Labor-Force Participation Rates and the Supply of Labor. Journal of Political Economy. V. 81. N.3.

9) BEN-PORATH, Yovan & WELCH, Finis (1976). Do sex preferences really matter? Quarterly Journal of Economics. V. 90, N. 2, 285-307.

10) BROWING, M. (1992). Children and Household Economic Behavior. Journal of Economic Literature, 30, 1434-1475.

11) CAMPÊLO, Ana Katarina & SILVA, Everton Nunes (2002). Children and family income: instrumental variables estimation of quantile treatment effects. Anais do XXIV Encontro da Sociedade Brasileira de Econometria.

12) COLEMAN, M. T. & PENCAVEL, J. (1993). Trends in market behavior of women since 1940. Industrial and Labor Relations Review, 46, 653-676.

13) De TRAY, Dennis (1973). Child quality and the demand for children. Journal of Political Economy. V. 81, N. 2, S70-S98.

14) FRÖLICH, Markus (2002). Nonparametric IV estimation of local average treatment effects with covariates. Department of Economics University of St. Gallen: Discussion paper 2002-19.

15) GREENE, Willian H (2000). Econometric analysis. Fourth Edition, New York University.

16) GOLDIN, C. (1995). Career and family: college women look to the past. NBER Working Paper N. 5188, Julho.

17) GONZAGA, Gustavo & SOARES, Rodrigo Reis (1999). Determinação de salários no Brasil: dualidade ou não-linearidade no retorno à educação. Revista de Econometria, Rio de Janeiro. V. 19, N. 2, Novembro.

18) GRONAU, Reuben (1973). The effect of children on the Housewife's value of time. Journal of Political Economy, V.81, N.2/ parte 2, Março/Abril.

19) GRONAU, Reuben (1977). Leisure, home production and work – the theory of allocation of time revisited. Journal of Political Economy. V. 84, N. 6, 1099-1124.

20) GRONAU, Reuben (1988). Sex-related wage differentials and women's interrupted careers – the chicken or the egg. Journal of Labor Economics, V. 6 N. 3, 277-301.

21) HECKMAN, James J. & MACURDY, Thomas E. (1980). A life-cycle model of female labor supply. Review of Economic Studies, V. 47 N. 1, 47-74.

22) HECKMAN, James (1995). Instrumental variables: a cautionary tale. National Bureau of Economic Research: Technical working paper No. 18

23) IMBENS, G.W., & ANGRIST, J.D. (1994). Identification and estimation of local average treatment effects. Econometrica, 62, 467-476.

24) INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA, (IBGE). www.ibge.org.br

25) KOENKER, R., & BASSETT, G. (1978). Regression Quantiles. Econometrica, 46, 33-50.

26) KOENKER, R. & PORTNOY, S. (1996). Quantile Regression. University of Illinois at Urbana-Champaign. Working Paper, N. 97-0100.

27) KOENKER, R. (2000). Galton, Edgeworth, Frisch and prospects for quantile regression in econometrics. Journal of Econometrics, 95. 347-374.

28) KORENMAN, S. & NEUMARK, D. (1992). Marriage, Motherhood and wages. Journal of Human Resources, 27, 233-255.

29) LEIBENSTEIN, Harvey (1957). Economic backwardness and economic growth: studies in the theory of economic development. New York: Wiley.

30) MACIEL, Marinalva Cardoso. A dinâmica das mudanças na distribuição salarial e no retorno à educação para mulheres: uma aplicação de regressão quantílica. Dissertação de Mestrado – Departamento de Estatística – UFPE, 2001.

31) POWELL, J. L.(1986). Censored regression quantiles. Journal Econometrics, 32, 143-155.

32) RIBEIRO, Eduardo Pontual (1997). Conditional labor supply quantile estimates in Brazil. Universidade Federal do Rio Grande do Sul: Texto para discussão N. 97/02.

33) RUBIN, D.B (1978). Bayesian inference for causal effects: The role randomization. Annals of statistics, 6, 34-58.

34) SCHULTZ, Theodore W (1973). The value of children: An Economic Perspective. Journal of Political Economy. V.81, N.2/ parte 2, Março/Abril.

35) SPANOS, Aris (1999). Probability theory and statistical inference: econometric modeling with observational data. Cambrigde University Press.

36) VELOSO, Fernando A (2000). Income composition, endogenous fertility and schooling investments in children. Anais do 22° Encontro de Econometria. Julho/2000.

37) WILLIS, Robert J. (1987). What have we learned from the Economics of the family? The American Economic Review. V.77, N.2, Maio

38) WILLIS, Robert J. (1973). A new approach to the Economic Theory of fertility Behavior. Journal of Political Economy. V.81, N.2, parte 2, Março/Abril.