

Hugo Borges Jales

*Peer Effects* na Educação no Brasil.  
Evidência a partir dos dados do SAEB

São Paulo  
2010



Hugo Borges Jales

*Peer Effects* na Educação no Brasil.  
Evidência a partir dos dados do SAEB

Dissertação apresentada à Escola de  
Economia de São Paulo, para a obtenção  
de Título de Mestre em Economia

Orientador: Sergio Pinheiro Firpo

São Paulo  
2010

Jales, Hugo B.

*Peer Effects* na Educação no Brasil. Evidência a partir dos dados do SAEB / Hugo Borges Jales. - 2010

68 páginas

Dissertação (Mestrado) - Fundação Getulio Vargas. Escola de Economia de São Paulo - EESP-FGV.

1. Educação
2. *Peer Effects*
3. *Social Interactions*

I. Fundação Getulio Vargas. Escola de Economia de São Paulo.

## Comissão Julgadora:

---

Prof. Dr.  
Vladimir Pinheiro Ponczek

---

Profa. Dra.  
Cristine Campos de Xavier Pinto

---

Prof. Dr.  
Sergio Pinheiro Firpo

*Dedicatória...*  
À minha esposa, Marina.

## Epígrafe

95 % das estatísticas estão erradas.

Homer J. Simpson

## Agradecimentos

A CAPES, pelo financiamento desta pesquisa. Aos meus amigos e familiares, pelo apoio. Aos colegas de turma, cuja amizade, companhia e camaradagem tornaram mais leve o difícil trabalho de completar este curso. Aos professores e funcionários da ESSP-FGV, cujos esforços diários fazem desta escola o centro de excelência que é.

Ao professor Sergio, por ter ido muito além do que se espera de um orientador. Pela paciência, boa vontade e amizade.

À minha esposa Marina, por tudo.

## *Resumo*

Este trabalho estima o efeito das Interações Sociais no desempenho dos estudantes no Brasil, a partir dos dados do SAEB 2005. Os resultados apontam que *Peer Effects* explicam uma parte importante da proficiência observada no exame de Matemática. Além disto, encontramos evidências de que este efeito é concentrado nos alunos no topo da distribuição de habilidade, o que sugere que políticas de *ability tracking* podem ser ótimas.

**Palavras-chave:** Educação, *Social Interactions*, *Peer Effects*



### *Abstract*

In this paper, I estimate the size of social interactions in the Brazilian educational production function, from the data called “SAEB-2005”. The results show that peer effects are important determinant of observed achievement in the Math exam. Furthermore, I found evidence that this effect is concentrated in the top of the ability distribution. This suggests that policies like “ability tracking” may be optimal.

**Keywords:** Education, Social Interactions, Peer Effects

# Lista de Figuras

4.1	Relação entre Pré-Escola e Proficiência . . . . .	33
4.2	Estimação de <i>peer effects</i> na presença de <i>outliers</i> . . . . .	45

# Lista de Tabelas

4.1	Estatísticas Descritivas - SAEB 2005 . . . . .	26
4.2	Estatísticas Descritivas - Regressões Univariadas . . . . .	27
4.3	Primeiro Estágio - Regressão <i>Within</i> . . . . .	28
4.4	Segundo Estágio - Regressão <i>Between</i> . . . . .	30
4.5	<i>Peer Effects</i> - Modelo Linear . . . . .	31
4.6	Efeitos ao Longo da Distribuição $\tau = 0.50$ . . . . .	36
4.7	Efeitos ao Longo da Distribuição $\tau = 0.10$ . . . . .	37
4.8	Efeitos ao Longo da Distribuição $\tau = 0.90$ . . . . .	38
4.9	<i>Peer Effects</i> - Modelo IV1 . . . . .	41
4.10	<i>Peer Effects</i> - Modelo IV2 . . . . .	43
4.11	Análise de Influência - Modelo Linear . . . . .	46
4.12	Análise de Influência - $\tau = 0.50$ . . . . .	47
4.13	Análise de Influência - $\tau = 0.10$ . . . . .	47
4.14	Análise de Influência - $\tau = 0.90$ . . . . .	48
4.15	<i>Peer Effects</i> - Modelo Estendido . . . . .	50
4.16	Modelo Estendido - $\tau = 0.10$ . . . . .	52
4.17	Modelo Estendido - $\tau = 0.50$ . . . . .	53
4.18	Modelo Estendido - $\tau = 0.90$ . . . . .	54
4.19	Análise de Influência - Modelo Estendido . . . . .	57
4.20	Análise de Influência - Modelo Estendido - $\tau = 0.10$ . . . . .	58

4.21	Análise de Influência - Modelo Estendido - $\tau = 0.50$ . . . . .	59
4.22	Análise de Influência - Modelo Estendido - $\tau = 0.90$ . . . . .	60

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	A relação entre <i>Inputs</i> escolares e desempenho . . . . .	1
1.2	<i>Peer Effects</i> . . . . .	3
<b>2</b>	<b>Interações Sociais</b>	<b>5</b>
2.1	Introdução . . . . .	5
2.2	Interações Sociais na Educação . . . . .	7
<b>3</b>	<b>Modelo</b>	<b>12</b>
3.1	Hipóteses e Identificação do Modelo . . . . .	18
<b>4</b>	<b>Resultados</b>	<b>25</b>
4.1	Dados e Estatísticas Descritivas . . . . .	25
4.2	Efeitos ao Longo da Distribuição . . . . .	35
4.3	Robustez . . . . .	40
4.4	Análise de Influência . . . . .	44
4.5	Extensões . . . . .	49
<b>5</b>	<b>Considerações Finais</b>	<b>61</b>
	<b>Referências Bibliográficas</b>	<b>63</b>

<b>A Primeiro apêndice</b>	<b>67</b>
A.1 Construção dos Indicadores . . . . .	67

# Capítulo 1

## Introdução

### 1.1 A relação entre *Inputs* escolares e desempenho

Os economistas reconhecem há tempos a importância da educação para a sociedade. Isto pode ser notado nos trabalhos de Mincer (1958), Lucas (1988), Ljungqvist (1993), dentre outros. Além do conhecido efeito positivo na remuneração do trabalho, advinda dos ganhos de produtividade, as diferenças na qualidade da educação entre os países costumam ser invocadas para explicar diferenças nos níveis de saúde, criminalidade, participação política, entre outros fatores (HANUSHEK, 2007).

A idéia de que os ganhos sociais advindos de uma melhoria na educação superem os ganhos privados não é exatamente nova. De fato, mesmo nos primórdios do pensamento econômico a formação do capital humano já era considerada de suma importância para o desenvolvimento econômico, crescimento e melhoria na distribuição da renda. Marshall (1890), no sexto capítulo do “Princípios de Economia”, ressalta que “nenhuma mudança seria mais conducente a um rápido aumento da riqueza nacional como uma melhoria das escolas”.

Embora exista um relativo consenso com relação à importância da educação na sociedade, ainda pouco se sabe sobre a melhor maneira de se promover melhorias no

ensino. Desde a publicação do “relatório Coleman” (Coleman et. al., 1966), os economistas buscam a compreensão de como as diversas características – dos alunos, das escolas e das famílias – se combinam na determinação do desempenho dos estudantes. A compreensão deste mecanismo, isto é, a mensuração da “função de produção de educação”, é crucial para a formulação de políticas públicas. Como ressaltado por Sacerdote (2001), *“it is clearly difficult to think about improving student outcomes in primary and secondary schools until we know which inputs matter”*.

No Brasil, a manutenção da educação básica é uma atividade que envolve a utilização de quantidades consideráveis dos recursos disponíveis no país. Isto pode ser ilustrado pela receita de 83 bilhões de reais do Fundo de Manutenção e Desenvolvimento da Educação Básica no ano de 2010.

Apesar do esforço para a universalização do acesso ao ensino básico e fundamental realizado nas últimas décadas e do volume de recursos despendido anualmente nesta atividade, a educação no país ainda apresenta sérias deficiências. Segundo a avaliação realizada pela Unesco, o Brasil figura apenas na 88ª colocação entre os países cujo ensino foi avaliado, atrás de países como Paraguai (72°), Bolívia (79°) e Colômbia (75°). Os resultados obtidos na avaliação realizada pelo programa PISA – Programme for International Student Assessment – apontam na mesma direção. Dentre as 57 nações avaliadas no ano de 2006, o Brasil figura entre os 7 países com a pior proficiência em matemática e leitura. Resolver a questão da qualidade do ensino no país é, portanto, um sério desafio para os formuladores de políticas públicas.

A despeito da sua importância, a estimação da função de produção educacional no Brasil foi por muitos anos uma tarefa impossível. A razão para isto é que somente encontravam-se disponíveis informações sobre os anos de escolaridade – o que, conforme ressaltado por Hanushek (2007) é uma medida extremamente imprecisa das habilidades cognitivas de um indivíduo –. Recentemente, com a implementação pelo Ministério da Educação de exames padronizados, a tarefa de compreender os determinantes do



desempenho educacional dos estudantes tornou-se mais factível (ALBERNAZ et. al. 2002). Nesta literatura, destacam-se os trabalhos de Barbosa e Fernandes (2001) e Albernaz et. al. (2002).

## 1.2 *Peer Effects*

Recentemente, a literatura internacional tem direcionado atenção para um particular fenômeno do processo de aprendizagem, conhecido como *peer effect*. Grosso modo, podemos definir *peer effect* como a influência dos membros de um grupo no desempenho dos seus colegas.

*Peer effect* surge naturalmente no ambiente da sala de aula. Por compartilharem o ambiente de aprendizado, as ações de um indivíduo podem interferir no desempenho dos demais. Isto ocorre, por exemplo, quando um aluno toma a decisão de conversar durante a exposição.

Se isto ocorre, isto é, se existirem *peer effects*, o desempenho esperado de um aluno passa a ser função de características dos seus colegas. Sendo assim, é necessário que se conheça a magnitude deste fenômeno para que se responda a importantes questões como os efeitos de políticas de *ability tracking* ou qual a forma ótima de se alocar os alunos entre as turmas. Como se pode notar, a forma com que se alocam os alunos pode se tornar um instrumento de política educacional capaz de alterar o desempenho esperado dos alunos. Além de ser necessária para a formulação de políticas educacionais, a importância dos *peer effects* também se deve ao fato de que na sua presença a função de produção educacional passa a dotar de uma singular característica de amplificação de variações iniciais, conhecida na literatura como *social multiplier*.

É possível encontrar alguma evidência a respeito da importância de características dos pares no desempenho dos alunos nos trabalhos de Fletcher (1997), Albernaz et. al. (2002) e Franco et. al. (2004). A literatura nacional, no entanto, ainda carece

de investigações mais detalhadas a respeito deste fenômeno, uma vez que exceto pelo trabalho de Pinto (2008), é praticamente inexistente no país pesquisas cujo foco resida na identificação e estimação de *peer effects*.

Sendo assim, este trabalho terá como objetivo lançar um pouco mais de luz a respeito deste fenômeno. Isto é, dentre os diversos fatores que interferem no desempenho dos estudantes, iremos buscar evidência a respeito da existência e da magnitude do efeito deste insumo particular, relacionado às características do grupo no qual cada aluno está inserido. Em outras palavras, faremos uma estimação da função de produção educacional que permite a presença de *peer effects* a partir das informações presentes nos dados de exames padronizados do Brasil.

## Capítulo 2

# Interações Sociais

*While economics has long focused on how individual decisions are interconnected via markets, there has for the last decade or so developed growing interest in understanding how social factors beyond the marketplace affect individual decisions and outcomes. Economic analysis now incorporates a range of dimensions in which individuals interact directly with one another, rather than indirectly via the effects of individuals on market prices. (...) The emergence of the social interactions literature parallels the rise of game theory, in which the key primitive assumptions are based on modeling how the behaviors of others affect an individual (...). Such direct interdependences in behaviors and outcomes are known in the economics literature as social interactions.*

Blume e Durlauf (2005)

### 2.1 Introdução

A dependência direta dos comportamentos e características dos indivíduos no comportamento dos demais foi um fenômeno relativamente ignorado pelos economistas até

recentemente. Uma das explicações para isto é o fato da economia tradicional ser fundamentada metodologicamente em decisões atomizadas de agentes individualistas (AKERLOF, 1997).

Esta forma direta de dependência, este tipo específico de externalidade, ficou conhecido na literatura como Interações Sociais. Um outro nome para este fenômeno é *non-market interactions*<sup>1</sup>, o que salienta a diferença de canais entre as Interações Sociais e o tradicional mecanismo de preços (SCHEINKMAN, 2006).

Existe uma série de fenômenos em que Interações Sociais parecem ser parte importante da explicação. Cientistas Sociais há tempos reconhecem o papel desta interdependência em questões como criminalidade, consumo de álcool ou drogas ilícitas, gravidez precoce, *crashes* em bolsas de valores, entre outras.

Uma importante característica destes modelos é a capacidade de explicar fenômenos em que há grande variação nos resultados observados entre diferentes grupos sem que haja uma correspondente grande diferença nas características destes grupos. Isto ocorre porque, na presença de complementaridades estratégicas, uma determinada variação nos fundamentos implica em uma revisão da estratégia de todos os membros do grupo na mesma direção do choque inicial. Deste modo, o efeito total é o resultado da combinação do efeito do choque inicial acrescido do efeito indireto, a alteração do comportamento dos membros daquele grupo. A este efeito dá-se o nome de *Social Multiplier* (SCHEINKMAN, 2006).

Em modelos onde há Interações Sociais muitas vezes ocorre do equilíbrio em que as decisões são tomadas descentralizadamente ser distante do ótimo social. Esta característica não é de se surpreender, uma vez que Interações Sociais são um tipo particular de externalidade e, em geral, na presença de externalidades não é incomum o equilíbrio descentralizado ser sub-ótimo. Sendo assim, pode ser possível a construção de políticas

---

<sup>1</sup>Uma série de outros termos utilizados na literatura também remetem à idéia de *Peer Effects*. São exemplos “*social norms*”, “*peer influences*”, “*neighbourhood effects*”, “*conformity*”, “*epidemics*”, “*interdependent preferences*”, entre outros (Manski, 1993).

públicas ou outros mecanismos de coordenação que promovam ganhos de Pareto por meio da internalização destes efeitos indiretos na tomada de decisão dos agentes.

## 2.2 Interações Sociais na Educação

*Human learning presupposes (...) a process by which children grow into the intellectual life of those around them.*

*We propose that an essential feature of learning is that it (...) awakens a variety of internal developmental processes that are able to operate only when the child is interacting with people in his environment and in cooperation with his peers.*

Lev S. Vygotsky

No campo da economia da educação, as *social interactions* são também denominadas de *peer effects*.

Educadores há muito reconhecem a interdependência presente no aprendizado infantil. Desde os trabalhos do psicólogo russo Lev. Vygotsky (1896-1934), o processo de aprendizado infantil é visto em parte como algo social, dependente do meio em que a criança se encontra.

Crianças passam boa parte de seu tempo interagindo com seus colegas, seja dentro ou fora da sala de aula. Neste processo de interação, as crianças se influenciam mutuamente. Estas interações podem afetar o aprendizado dos alunos. Exemplos simples deste mecanismo são o empréstimo de materiais, trabalhos em grupo, etc. Também a disponibilidade de livros para serem compartilhados em uma turma depende da disponibilidade de livros dos colegas. De forma análoga, a qualidade dos trabalhos em grupo e, portanto, do aprendizado ocorrido neste processo, dependerá da qualidade do grupo como um todo. Além disto, colegas podem efetivamente ajudar (ou “ensinar”) uns aos outros.

Peer effects podem também surgir através de algumas decisões tomadas pelos alunos em sala de aula, como realizar uma boa pergunta ou conversar durante a exposição. Neste caso, pode se pensar a atenção do professor e ou silêncio na sala de aula como bens públicos que são compartilhados pelo grupo.

Uma última fonte interessante de *peer effects* são as pressões para comportamento conforme a norma. Muitas vezes, a capacidade de efetivamente se relacionar com os colegas esta relacionada com regras de comportamento internas a cada uma das turmas. Neste caso, ocorrem pressões para que os membros do grupo se comportem de forma semelhante. Por exemplo, um bom aluno pode reduzir seu esforço em uma turma onde ele não encontra colegas semelhantes para evitar o rótulo de “*nerd*”. Este mesmo aluno, em uma classe repleta de bons alunos, não teria incentivos para reduzir o seu esforço. Pelo contrário, talvez seu status nesta nova turma dependesse fundamentalmente da sua capacidade de demonstrar habilidades tão boas quanto as dos seus pares <sup>2</sup> (COOLEY, 2007).

Por todas as razões apresentadas acima, educadores costumam crer que a composição da turma interfere no desempenho do aluno – além, obviamente, das características inerentes ao indivíduo e à escola. Como vimos, isto pode ocorrer por motivos “psicológicos” como, por exemplo, alunos tenderem a se comportar de acordo com regras sociais internas ao grupo ou mesmo porque os estudantes geram externalidades aos seus colegas de acordo com o seu comportamento em sala.

Coexistem na literatura teórica educacional diversas hipóteses a respeito da forma ótima de se alocar estudantes dotados de diferentes níveis de habilidade e aprendizado pretérito. Como exemplo, podemos citar a conhecida por “*one shining star*”, que significa que alunos de alta performance beneficiam os demais em uma medida maior do que são prejudicados por eles.

Na direção oposta, a hipótese conhecida por “*the bad apple*” supõe que os alunos de

---

<sup>2</sup>Sobre isto, ver também Bishop et. al. (2003).

pior rendimento afetam negativamente os demais de forma mais intensa que são beneficiados por eles. Ainda existem as hipóteses de que os alunos beneficiam-se da existência de um ambiente diverso (*rainbow*), ou pelo contrário, de um ambiente homogêneo, onde o professor tem mais facilidade de ministrar o conteúdo (*boutique*). Através identificação da heterogeneidade dos *peer effects* ao longo da distribuição de desempenho, podemos verificar a plausibilidade de cada uma destas hipóteses ao caso brasileiro (HOXBY e WEINGARTH, 2006).

Seja a origem deste efeito de natureza psicológica, associada à pressão pela conformidade com a norma no grupo; tecnológica, referente aos ganhos de escopo quando o professor ministra o conteúdo em grupos mais homogêneos; ou econômica, associada às externalidades inerentes ao compartilhamento do ambiente de estudo, espera-se que o desempenho do aluno seja relacionado a características não apenas suas mas, também, dos seus colegas.

Este fato tem importantes implicações para políticas públicas na educação. Primeiramente, porque realocações de estudantes, sejam dentro de uma mesma escola ou entre escolas de uma mesma vizinhança, são virtualmente sem custo. Isto significa que é possível promover melhorias no desempenho sem que seja necessário o aumento nos recursos destinados à educação. Além disto, realocações dos pares podem promover não apenas ganhos no desempenho esperado mais também alterações na sua variabilidade. Deste modo, a depender da magnitude dos *peer effects* ao longo da distribuição de desempenho pode ser formado um *trade-off* entre equidade e eficiência nas políticas de realocação de alunos.

Apesar de seu interessante apelo teórico, existe uma série de dificuldades empíricas envolvidas na mensuração da magnitude dos efeitos de Interações Sociais. Em razão disto, a evidência empírica a respeito deste tema é, na maioria dos casos, longe de ser conclusiva (DURLAUF, 2002).

De fato, sobre diversas condições, a mensuração da existência destes efeitos é sequer

possível. Este problema, conhecido na literatura como “*Reflection Problem*” (Manski, 1993), decorre da dificuldade em se isolar efeitos de variáveis que caminham quase sempre em conjunto: os comportamentos e as características dos grupos. Além disto, uma outra dificuldade consiste na separação dos efeitos de interações sociais dos efeitos de inputs não observados compartilhados dentro do grupo. Em razão disto, existe um extenso ramo da literatura de interações sociais trata das condições sobre as quais os parâmetros destes modelos são identificáveis.

Diante destas dificuldades, diversos métodos foram utilizados no intuito de estimar a extensão dos *peer effects* na educação e, até o momento, não existe consenso na literatura sobre quais técnicas são mais adequadas para se estimar esta relação. Como exemplo de trabalho relacionado à estimação de *peer effects* podemos citar Hoxby (2000), que explora variações de gênero e raça entre coortes adjacentes como forma de controlar para o processo de auto-seleção dos grupos. Lavy et. al. (2009) utilizam-se dos desempenhos em exames padronizados realizados em diferentes disciplinas e pontos do tempo como medidas de habilidade dos pares. Sacerdote (2000) faz uso de um banco de dados experimental onde a alocação dos alunos nos dormitórios foi decidida de forma aleatória. Ammermueller e Pischke (2006) exploram variações entre turmas dentro de uma mesma escola.

Este trabalho se relaciona mais proximamente com o realizado por Cooley (2007) e Pinto (2008). O primeiro, no contexto dos determinantes de desempenho dos alunos do estado norte-americano da Carolina do Norte utiliza de dados em painel de modo a controlar para o viés associado à habilidade não-observada do professor. Finalmente, Pinto (2008) estima para a mesma população de interesse que a deste trabalho, isto é, alunos da quarta série do ensino fundamental no Brasil, uma função de produção com a presença de *peer effects* utilizando da regra de alocação dos alunos como forma de contornar a auto seleção na formação dos grupos. A principal diferença entre este trabalho e o realizado por Pinto (2008) consiste na estimação conjunta de efeitos contextuais e



*peer effects* endógenos.

## Capítulo 3

# Modelo

*The two rival explanations for excess between-group variance, group-level heterogeneity and social interactions, are straightforward to understand, but exceptionally difficult to discriminate between empirically. Hoxby (2002, p. 58) emphasizes the “formidable obstacles” faced by researchers when attempting to detect peer effects in the learning process. In a recent and wide-ranging review Durlauf (2002, p. 20) concludes that “there is little reason why a skeptic should be persuaded to change his mind by the statistical evidence [on social interactions] currently available”. Often associated with controversy, the empirical literature on social interactions is also characterized by widely divergent conclusions across different researchers.*

Graham (2005)

O modelo mais utilizado na literatura para a estimação de *peer effects* é chamado de *linear-in-means*. São duas as simplificações características deste modelo: o desempenho de um aluno é função apenas da média das características dos seus colegas e a relação entre o desempenho e as demais covariadas é linear.

Nesta seção iremos apresentar um modelo microeconômico simples de escolhas na

presença de interações sociais, baseado nos trabalhos de Graham (2005) e Akerlof (1997), cujas condições de equilíbrio resultam em um modelo *linear-in-means*.

Seja  $y_{ig}$  o desempenho esperado do indivíduo  $i$  da turma  $g$ ,  $X_{ig}$  o vetor composto pelas características do indivíduo, assuma que sejam observados  $g = 1, 2, 3, \dots, G$  grupos com um número possivelmente distinto de componentes em cada um. Dadas as informações sobre o indivíduo, a escola e o desempenho esperado dos indivíduos dentro dos grupos ( $m^e$ ); os alunos fazem uma escolha de desempenho que maximiza a seguinte função de utilidade indireta <sup>1</sup>:

$$\nu(y_{ig}|k(x, g, i)) = -\frac{1 - \beta_3}{2}y_{ig}^2 + k(x, i, g)y_{ig} + \frac{\beta_3}{2}(y_{ig} - m_{ig}^e)^2$$

Os dois primeiros termos desta função correspondem ao que Akerlof denomina de utilidade intrínseca, enquanto o último corresponde ao componente de interação social <sup>2</sup>. O formato do último termo implica que os indivíduos apresentam desutilidade ao se distanciarem da norma dentro de seu grupo, gerando o comportamento “conformista”. Considerando que os indivíduos fazem a escolha do nível de desempenho ótimo considerando como fixo o comportamento do grupo, temos que o desempenho ótimo é dado por:

$$y_{ig}^{BR} = k(x, g, i) + \beta_3 m_{ig}^e$$

A função  $y_{ig}^{BR}$  mapeia das características do indivíduo e do comportamento esperado do grupo nas escolhas ótimas do indivíduo. Conforme ressaltado por Blume e

---

<sup>1</sup> O termo escolha pode parecer mal empregado. O que deve ser entendido aqui é que todo aluno tem a capacidade, via esforço, de determinar ainda que parcialmente o seu desempenho. Dadas as suas características, a variável de ajuste que os alunos dispõem é apenas o quanto irão se esforçar. O *trade-off* com o qual cada um se defronta é entre a utilidade de um desempenho maior e o custo de se esforçar ligeiramente mais. As diferenças nas características dos indivíduos se refletem em diferenças no desempenho por meio de diversos canais. Por exemplo, pais mais atenciosos podem levar a um desempenho maior por aumentar a sensação de satisfação ao obter um desempenho melhor; enquanto professores capazes podem aumentar o desempenho ao tornar o esforço necessário para o aprendizado menor (COOLEY, 2007).

<sup>2</sup>Note que o indivíduo apenas responderá ao comportamento médio no grupo quando  $\beta_3$  for diferente de zero. Assim, testar se  $\beta_3$  é igual a zero significa testar a presença de *peer effects* endógenos

Durlauf (2005), este modelo apenas restringe o comportamento dos indivíduos após impormos condições sobre como eles formam suas expectativas com relação ao comportamento do grupo. Note que, deixando livre como as expectativas são formadas, podemos conciliar com o modelo qualquer conjunto de *outcomes* observado por uma apropriada escolha de  $m_{ig}^e$ .

Diante disto, iremos impor neste modelo uma condição conhecida por *self-consistency*. Esta condição, usualmente utilizada em modelos de interação social, impõe que o comportamento esperado de um grupo, isto é,  $m_{ig}^e$ , é aquele obtido quando se toma a esperança dentro do grupo dos desempenhos ótimos a partir das funções de melhor resposta. Assim:

$$E[y_g] = m_{ig}^e$$

Tomando a esperança dentro do grupo e fazendo as substituições devidas, temos que:

$$E[y_{ig}] = \frac{1}{1 - \beta_3} E[k(x, i, g)]$$

Tal qual ressaltado por diversos autores, como Manski (1995), Angrist, Grady e Imbens (2000) e Graham (2005), esta hipótese tem como objetivo contornar o problema de *missing data* implícito na estimação dos parâmetros deste modelo <sup>3</sup>.

O parâmetro  $k$  deste problema varia entre os indivíduos, determinando o nível de desempenho esperado. Em princípio, a forma com que se  $k$  altera como função de  $x, i$ , e  $g$  pode ser totalmente flexível. Fazendo uma simplificação usual em econometria aplicada, vamos supor que  $k$  é uma função linear do vetor de covariadas  $X$ . Assim, o

---

<sup>3</sup>De forma análoga, a hipótese de equilíbrio é fundamental para que se resolva o problema de *missing data* existente na estimação de funções de oferta e demanda. Ver Graham(2005).

problema toma a seguinte forma <sup>4</sup>:

$$k(x, i, g) = \beta_0 + X'_{ig}\beta_1 + \overline{X}'_g\beta_2 + \alpha_g + \epsilon_{ig}$$

E a função que relaciona  $y$  com as características do indivíduo, do grupo e com o desempenho esperado dentro do grupo toma o formato conhecido na literatura como *linear-in-means* <sup>5</sup>:

$$y_{ig} = \beta_0 + X'_{ig}\beta_1 + \overline{X}'_g\beta_2 + \beta_3\overline{y}_g + \alpha_g + \epsilon_{ig}$$

Manski (1993), estabeleceu 2 resultados negativos com relação à identificação dos parâmetros do modelo *linear-in-means*. O primeiro deles diz que sobre hipóteses usuais sobre a distribuição dos erros, os parâmetros deste modelo não podem ser identificados a menos que se restrinja de alguma maneira a forma com que se manifestam os efeitos contextuais. Ou seja, é necessário que se imponha alguma restrição no vetor  $\beta_2$ . Vejamos porque isto ocorre:

Primeiramente, devemos notar que não podemos realizar inocentemente uma regressão dos dados na forma estrutural em que o modelo foi apresentado, uma vez que o coeficiente  $\hat{\beta}_3$  irá convergir para 1 a despeito do valor do parâmetro na população (GRAHAM, 2005).

Tomando a esperança de  $k$  e substituindo na equação original, temos a seguinte

---

<sup>4</sup>Alguns autores, como Graham e Hahn (2005), apresentam o modelo de modo que os indivíduos reagem às expectativas de comportamento e características dentro do grupo, em vez das médias. Esta distinção faz com que o termo de erro final incorpore além da variação idiossincrática um componente referente ao *sampling error* no cálculo das esperanças. Esta alteração, no entanto, não traz grandes mudanças nas principais características do modelo. Deste modo, iremos assumir que os indivíduos reagem à média das características e comportamento dentro do grupo.

<sup>5</sup>Por limitações inerentes à disponibilidade dos dados, usualmente se assume que o grupo em que o aluno se relaciona consiste apenas na turma no qual ele está inserido, e a influência de cada membro deste grupo no comportamento dos demais é igual para todos os alunos. Calvó-Armengol, Patacchini e Zenou (2008), utilizando de um banco de dados que contém a relação de proximidade de cada indivíduo com os demais, estimam a magnitude dos peer effects na educação primária nos EUA flexibilizando estas duas hipóteses.

forma reduzida:

$$y_{ig} = \beta_0 + X'_{ig}\beta_1 + \bar{X}'_g\beta_2 + \frac{\beta_3}{1-\beta_3}(\beta_0 + \bar{X}'_g(\beta_1 + \beta_2) + \alpha_g + \bar{\epsilon}_{ig}) + \alpha_g + \epsilon_{ig}$$

O que é equivalente a:

$$y_{ig} = \frac{\beta_0}{1-\beta_3} + X'_{ig}\beta_1 + \bar{X}'_g \frac{\beta_3\beta_1 + \beta_2}{1-\beta_3} + \frac{\bar{\epsilon}_g + \alpha_g}{1-\beta_3} + \epsilon_{ig} = \pi_0 + X'_{ig}\pi_1 + \bar{X}'_g\pi_2 + u_{ig}$$

Note que ainda que pudéssemos assumir formas convenientes para a distribuição dos termos de erro,  $\alpha_g$  e  $\epsilon_{ig}$ , de forma a permitir a identificação consistente dos parâmetros desta equação por OLS, ainda assim não atingiríamos o objetivo de identificar  $\beta_2$  e  $\beta_3$ . Isto ocorre porque o número de parâmetros obtidos a partir desta forma reduzida é insuficiente para recuperar os coeficientes da forma estrutural do modelo. Veja que:

$$\pi_0 = f(\beta_0, \beta_3)$$

$$\pi_1 = \beta_1$$

$$\pi_2 = f(\beta_3, \beta_1, \beta_2)$$

Isto significa que apenas  $\beta_1$ , que está associado à influência das características individuais pode, sob as condições até então apresentadas, ser identificado. Isto ocorre porque pode sempre ser escrito como uma combinação linear das demais covariadas do modelo. Em outras palavras, é difícil o processo de separação do que é o efeito de uma alteração na característica do grupo –  $\bar{X}_g$  – do que é um efeito de uma alteração no comportamento do grupo –  $\bar{y}_g$  – uma vez que estas duas variáveis andam sempre perfeitamente em conjunto. A este fenômeno deu-se o nome de *Reflection Problem* (MANSKI, 1993). Sobre isto, o autor ressalta que:

*“The “reflection” problem arises when a researcher observing the distribution of behaviour in a population tries to infer whether the average behaviour in some group influences the behaviour of the individuals that comprise the group. The term reflection is appropriate because the problem is similar to that of interpreting the almost simultaneous movements of a person and his reflection in a mirror. Does the mirror cause the person’s movement or reflect them? An observer who does not understand something of optics and human behaviour would not be able to tell.”*

Manski (1993)

Portanto, para que o modelo seja identificado, é necessário que o vetor de parâmetros estimados na forma reduzida seja, no mínimo, igual ao vetor de parâmetros da forma estrutural. Isto pode ser obtido através da adição de mais uma equação no sistema. Em geral, a equação utilizada é a restrição é de que um dos elementos do vetor  $\beta_2$  é igual a zero.

Finalmente, Manski (1993) também ressaltou a dificuldade em se separar o que são efeitos das características ou comportamentos do grupo de inputs não observados que compartilhados pelos seus membros, como a habilidade do professor. Como a formação de grupos não é um processo aleatório, é muitas vezes difícil argumentar se os dados indicam a presença de interações sociais ou de um forte componente de auto-seleção. Esta distinção é de particular importância porque as diferentes justificativas para o excesso de variação de desempenho entre os grupos, *peer effects* e heterogeneidade no nível do grupo, trazem consigo diferentes implicações de políticas.

Como os parâmetros de interesse deste modelo  $\beta_2$  e  $\beta_3$  variam no nível do grupo, não é possível utilizar da metodologia de efeito fixo para estimar os parâmetros de forma robusta a presença de variáveis não observadas no grupo correlacionadas com o vetor de covariadas. Este problema é análogo ao estudado por Hausmann e Taylor (1981), no

contexto da estimação com dados em painel do efeito marginal de variáveis que variam apenas entre indivíduos na presença de heterogeneidade individual não observada.

Blume e Durlaufi (2005) ressaltam que, se assumirmos que  $\alpha$  pode ser arbitrariamente correlacionado com  $X$ , qualquer vetor de parâmetros associados tal que  $\beta_2$  e  $\beta_3$  são diferentes de zero, isto é, consistentes com a hipótese de existência de *peer effects*, pode ser substituído por um vetor de parâmetros tal que os coeficientes  $\beta_2$  e  $\beta_3$  são nulos por meio de uma apropriada alteração no vetor  $\alpha$ , sem qualquer alteração no ajuste do modelo. Deste modo, a identificação de *peer effects* na presença de heterogeneidade não observada ao nível do grupo é uma tarefa não trivial.

Em geral, a solução para este problema consiste em encontrar variáveis instrumentais que apresentem as condições usuais de identificação, isto é, uma vez que se condicione nelas a distribuição dos erros no nível do grupo passa a ser não correlacionada com as demais variáveis do modelo. Outro caminho é encontrar condições tais que seja plausível supor que a distribuição dos erros,  $\alpha_g$  e  $\epsilon_{ig}$  sejam não correlacionadas com as demais variáveis do modelo. Finalmente também é possível o uso de diferentes estratégias, tais a exploração de dados em painel (de indivíduos e grupos ao longo do tempo <sup>6</sup>), ou a modelagem conjunta tanto do comportamento dos indivíduos quanto do processo de formação dos grupos.

Em razão de todos estes fatores, não é de se surpreender que a evidência empírica em favor da existência de *peer effects* seja controversa e inconclusiva, conforme ressaltado na citação no início desta seção.

### 3.1 Hipóteses e Identificação do Modelo

Nesta seção iremos apresentar e fazer uma breve discussão a respeito das hipóteses que utilizaremos para a identificação do modelo. Conforme discutido na seção anterior, a

---

<sup>6</sup> Ver Hoxby, 2000.



identificação de efeitos contextuais, isto é, *peer effects* relacionados às características dos pares, conjuntamente com o parâmetro relacionado com o desempenho dos pares (ou *endogenous peer effects*) depende de uma restrição de exclusão. Assim, a primeira hipótese para a identificação dos parâmetros deste modelo é:

H1: (*Exclusion Restriction*) A variável pré-escola não apresenta efeito contextual. Isto é, a proporção de alunos com pré-escola em uma classe não apresenta efeito direto no desempenho dos alunos, exceto aquele que ocorre, endogenamente, através do efeito da pré-escola no desempenho de cada um dos pares. Matematicamente, impomos que o elemento referente à pré-escola no vetor  $\beta_2$  é igual a zero.

Ter cursado a pré-escola é uma atividade que conhecidamente causa benefícios no desempenho dos alunos mesmo em séries mais avançadas. A proporção de alunos que cursaram pré-escola dentro de uma classe, no entanto, dificilmente poderia apresentar um efeito direto. É possível pensar em diversos canais pelos quais gênero, raça, escolaridade dos pais pode interferir diretamente no desempenho dos demais alunos. Por exemplo, pais mais educados podem auxiliar trabalhos em conjunto realizados em casa. Além disto, através de contato direto, estes podem incentivar os colegas dos seus filhos ao estudo, dentre muitas outras formas. Ter cursado pré-escola, por outro lado, é de difícil compreensão como poderia alterar o desempenho dos colegas. Uma possibilidade seria se houvesse alguma complementaridade nos recursos utilizados como, por exemplo, o material didático. Como isto não ocorre, consideramos razoável impor esta condição <sup>7</sup>.

Diversos autores sugerem a utilização de variáveis defasadas para a quebra da simultaneidade presente no modelo. Lavy et.al. (2009), em um contexto semelhante, faz uso da proficiência em exames passados para mensurar o efeito da habilidade dos pares. A hipótese que fazemos aqui é, em certo sentido, semelhante. Estamos impondo que ter

---

<sup>7</sup> Ainda que não concorde com a restrição de exclusão utilizada, um cético pode obter importantes informações a respeito da magnitude dos *peer effects* através dos resultados apresentados. Para isto, basta que se recalcule os parâmetros da forma estrutural substituindo o coeficiente que restringimos a ser zero por qualquer outro valor. Um exemplo desta estratégia seria impor que o efeito contextual da pré-escola não pode ser maior do que um arbitrário percentual do efeito desta variável no desempenho do indivíduo.

cursado a pré-escola é a pré-escola promove apenas um *shift* na habilidade (ou desempenho) do indivíduo e, principalmente, a proporção de alunos que cursou a pré-escola em uma classe não afeta de forma direta, exceto por meio do diferenciado desempenho destes alunos, o desempenho dos demais.

Assim, de modo a explicitar que esta variável não apresenta efeito contextual, iremos sempre escrevê-la separadamente das demais variáveis do modelo. Além disto, iremos separar o termo  $\alpha_g$  em sua parte observável e não observável. Assim:

$$y_{ig} = \beta_0 + X'_{ig}\beta_1 + \bar{X}'_g\beta_2 + \beta_3\bar{y}_g + r_{ig}\beta_4 + \kappa'_g\beta_5 + v_g + \epsilon_{ig}$$

Em que  $r_{ig}$  é um indicador igual a 1 caso o aluno  $i$  da turma  $g$  cursou a pré-escola, e  $\alpha_g = \kappa'_g\beta_5 + v_g$ . As demais hipóteses são referentes à distribuição conjunta das covariadas e dos termos não observados necessárias para a estimação consistente dos parâmetros. Seja  $Z_{ig} = (X'_{1g}, X'_{2g}, \dots, X'_{mg}, \bar{X}'_g, r_{1g}, \dots, r_{mg}, \kappa'_g)'$  e  $Z_g = (Z_{1g}, Z_{2g}, \dots, Z_{mg})'$  iremos impor que:

$$\text{H.2: (Exogenidade de } \epsilon_{ig} \text{ ) } E[\epsilon_{ig}|Z_g, \alpha_g] = 0$$

$$\text{H.3: (Random Effects) } E[v_g|Z_g] = E[v_g] = 0$$

Esta hipótese, H.3; é fundamental para a identificação do modelo. Através dela contornamos o problema de endogeneidade na formação dos grupos. Conforme já ressaltado nas seções anteriores, uma das dificuldades na obtenção de evidências críveis de *peer effects* reside no fato de que a formação dos grupos não é aleatória e, portanto, o que estamos chamando de *peer effects* pode simplesmente ser efeito de uma correlação entre características dos pares e inputs não observados no nível do grupo.

A principal preocupação com relação à auto-seleção reside numa provável correlação positiva entre o indicador socioeconômico e os determinantes não observáveis do desempenho no nível do grupo. Isto ocorre porque pais mais afortunados podem matricular seus filhos em escolas cuja qualidade dos professores é sabidamente superior (qualidade

esta dificilmente observada por nós econométristas). Este processo de auto-seleção nos levaria a encontrar evidência espúria em favor de efeitos contextuais relacionados ao indicador socioeconômico.

Nós contornamos este problema utilizando um vetor de controles no nível do grupo,  $\kappa_g$ , vasto o bastante para que seja crível que os determinantes no nível do grupo não observados de desempenho, após condicionar nas informações incluídas no modelo, não sejam correlacionados com o indicador socioeconômico. Para isto, nós utilizamos variáveis demográficas (se a escola se localiza em zona urbana, em capital), vinculação da escola (federal, estadual, municipal ou privada), características do professor e do diretor (raça, idade, sexo, escolaridade e experiência), além de indicadores de qualidade do material e instalações físicas e violência na escola.

Consideramos razoável assumir que o conjunto de informação disponível nesta especificação é praticamente tão bom quanto o disponível pelos pais quando matriculam seus filhos em uma determinada escola e, deste modo, as variações no desempenho entre as turmas, uma vez controlado por estas informações, não devem ser resultado de nenhum processo de auto-seleção, uma vez que somos capazes por controlar pela capacitação do corpo docente, violência, qualidade das instalações, qualidade da administração além das usuais variáveis demográficas<sup>8</sup>.

Talvez uma grande fonte de preocupação com esta estratégia seja a possibilidade do indicador socioeconômico ser correlacionado com a parte não observada da habilidade do professor (isto é, aquela fração da habilidade que não esta correlacionada com o sexo, raça, idade, escolaridade ou experiência do mesmo). Isto poderia ocorrer desde que o conjunto de informação dos pais incluísse a parte da habilidade do professor que não é disponível para nós econométristas.

Para contornar este problema, nós incluímos o salário do professor como *proxy* da

---

<sup>8</sup>Estamos cientes de que, tal como ressaltado por Heckman(1996, apud Blume e Durlauf, 2005) “*persons making decisions have more information about the outcomes than the statisticians studying them*”, mas tentamos reduzir o viés associado a esta “ignorância fundamental” a uma magnitude desprezível.

habilidade não observada. Assumido que a habilidade é observada pelo empregador e que há um mercado de trabalho, um resultado conhecido na microeconomia implica que o salário do professor será função da sua habilidade. Na presença de custos de transação, esta relação deixa de ser direta, uma vez que o professor somente altera de emprego caso o ganho esperado supere o custo de transação. Mas ainda assim a habilidade do professor impõe um *lower bound* no salário que ele aceita receber. Matematicamente, na presença de custos de mudança de emprego, podemos descrever de forma simplificada a relação entre salário e produtividade da seguinte forma:

$$w_i = \max[w^s, h_i - c]$$

Em que  $w^s$  é o salário real ofertado pelo empregador,  $h_i$  é a produtividade do professor e  $c$  é o custo de mudança de emprego. Se o empregador oferta um salário real muito pequeno, o professor decide pela mudança de emprego, paga o custo  $c$ , e recebe um salário igual à sua produtividade em outra escola. Caso o salário ofertado seja maior que a produtividade menos o custo de mudança, o professor permanece no emprego. Como se pode notar, a habilidade do professor impõe um limite inferior sobre o qual o salário não pode ser reduzido. Assim, controlando pelo salário do professor esperamos eliminar o viés advindo da impossibilidade de se observar a sua habilidade.

Deste modo, a inclusão desta variável ainda assim se justifica. Note que se esperamos que exista uma correlação entre a habilidade não observada e o indicador socioeconômico, esta correlação deve desaparecer quando condicionamos na informação do salário do professor, mesmo na presença de custos de transação. Isto ocorre porque a habilidade do professor e o indicador socioeconômico devem crescer conjuntamente de forma simples (como linearmente), enquanto o salário a habilidade devem crescer sempre acima de uma determinada reta de fronteira (pois o salário será sempre igual ou maior que o valor da habilidade menos o custo do professor trocar de emprego)<sup>9</sup>.

---

<sup>9</sup>Podemos enriquecer este modelo em diversas maneiras. Por exemplo, podemos introduzir incerteza

Assim, a inclusão da variável de salário como *proxy* da produtividade não observada do professor deverá controlar para a correlação entre habilidade não observada e indicador socioeconômico a menos que nosso erro de medida na habilidade do professor esteja correlacionado com o indicador socioeconômico, o que não parece fazer o menor sentido<sup>10</sup>. Portanto, parece razoável assumir que o salário do professor possui as características necessárias para funcionar como uma boa *proxy*<sup>11</sup> da habilidade do mesmo.

Assumindo as hipóteses apresentadas, podemos utilizar a metodologia proposta por Graham e Hahn (2005) para a estimação dos parâmetros do modelo. Explorando a variação de desempenho dentro do grupo, identifica-se  $\beta_1$  e  $\beta_4$ , isto é, o efeito das variáveis que apresentam variação entre os alunos de uma mesma turma.

Tomando os desvios em torno da média dentro de cada turma, temos que:

$$y_{ig} - \bar{y}_g = \ddot{y}_{ig} = [x_{ig} - \bar{x}_g]' \beta_1 + [r_{ig} - \bar{r}_g] \beta_4 + \ddot{\epsilon}_{ig}$$

Pela hipótese H.2,  $\ddot{\epsilon}_{ig}$  é não correlacionado com  $\ddot{X}_{ig}$  ou  $\ddot{r}_{ig}$ . Logo podemos estimar a equação acima por OLS, e recuperar os parâmetros  $\beta_1$  e  $\beta_4$

Tomando a média de desempenho dentro de cada uma das turmas, temos:

$$\bar{y}_g = \frac{\beta_0}{1 - \beta_3} + \bar{X}_g' \frac{\beta_1 + \beta_2}{1 - \beta_3} + \bar{r}_g' \frac{\beta_4}{1 - \beta_3} + \bar{\kappa}_g' \frac{\beta_5}{1 - \beta_3} + \frac{v_g}{1 - \beta_3} + \frac{\bar{\epsilon}_g}{1 - \beta_3} = \pi_1 + \bar{X}_g' \pi_2 + \bar{r}_g' \pi_3 + \bar{\kappa}_g' \pi_4 + \xi_g$$

---

no salário do professor ao trocar de emprego ou permitir heterogeneidade não apenas na habilidade mas também nos custos de mudança. Estas alterações, embora descrevam melhor a realidade, mantêm a essência do argumento, que é a correlação forte e positiva entre o salário do professor e a sua habilidade.

<sup>10</sup>Isto aconteceria, por exemplo, se os alunos com indicador socioeconômico maior tendessem a ser ensinados por professores não com habilidade maior, mas com a maior habilidade dentre aqueles que recebem o mesmo salário. Esta relação tem um sentido totalmente diferente do que usualmente se entende como o processo de auto-seleção na formação dos grupos, que é relacionado ao fato de que a qualidade do professor é um bem normal.

<sup>11</sup>Ver Wooldridge (2002).

Em que:

$$\xi_g = \frac{v_g + \bar{\epsilon}_{ig}}{1 - \beta_3}$$

Sobre as hipóteses apresentadas, podemos estimar os parâmetros  $\pi_1$ ,  $\pi_2$  e  $\pi_3$  através de OLS aplicados nas médias das variáveis nas turmas (*between regression*). Com as estimativas obtidas em ambas as regressões, álgebra simples permite recuperar os parâmetros da forma estrutural:

$$\hat{\beta}_3 = \frac{\hat{\pi}_3 - \hat{\beta}_4}{\hat{\pi}_3}$$

$$\hat{\beta}_2 = \hat{\pi}_2(1 - \hat{\beta}_3) - \hat{\beta}_1$$

Assim, podemos identificar os parâmetros do modelo *linear-in-means*. Na seção seguinte, apresentamos as estimativas destes parâmetros para a população de alunos da quarta série do ensino fundamental no Brasil, obtidas através dos dados do SAEB para a disciplina de matemática.

## Capítulo 4

# Resultados

### 4.1 Dados e Estatísticas Descritivas

Para a estimação da função de produção educacional no Brasil, utilizamos da base de dados conhecida como SAEB.

O SAEB - sigla cujo significado é Sistema de Avaliação Educacional Básica - tem como objetivo principal avaliar a qualidade da educação brasileira. Suas informações permitem a mensuração do desempenho do sistema educacional das diversas unidades da federação e servem para nortear a formulação eficiente de políticas públicas no âmbito educacional. A periodicidade desta pesquisa é bianual. Neste trabalho foram utilizadas as informações referentes ao desempenho no exame de proficiência em Matemática no exame do ano de 2005. As Tabelas [4.1](#) e [4.2](#) resumizam as principais características dos dados:

Tabela 4.1: Estatísticas Descritivas - SAEB 2005

Variável	Obs	Média	Erro-Padrão	Mínimo	Máximo
Proficiência Saeb	41783	188.90	48.94	65.43	373.44
Socioeconomic Index	27123	0.00	2.12	-4.23	6.07
Female	40964	0.50	0.50	0.00	1.00
White	39751	0.38	0.48	0.00	1.00
Parental Education	31103	6.04	2.64	1.00	9.00
Age	41170	10.66	1.20	8.00	15.00
Fail	40036	0.28	0.45	0.00	1.00
Preschool	39432	0.81	0.39	0.00	1.00
Violência Index	38074	0.00	1.37	-0.92	9.36
Quality School Index	41887	0.00	1.94	-7.30	1.50
Salário Professor	36953	905.35	628.38	0.00	3100.00
Dummy Urbano	43823	0.94	0.23	0.00	1.00
Dummy Capital	43823	0.48	0.50	0.00	1.00
Dummy Federal	43823	0.01	0.08	0.00	1.00
Dummy Estadual	43823	0.35	0.48	0.00	1.00
Dummy Municipal	43823	0.35	0.48	0.00	1.00
Dummy Privada	43823	0.30	0.46	0.00	1.00
<i>Teacher Characteristics</i>					
Gender	38209	0.93	0.26	0.00	1.00
Age	38079	39.11	9.76	20.00	60.00
Race	37878	0.46	0.50	0.00	1.00
Experience	37759	13.48	6.76	0.00	22.00
Schooling	36789	14.00	1.74	8.00	15.00
Grades Homework	41779	2.96	0.23	0.00	3.00
<i>Principal Characteristics</i>					
Gender	40555	0.81	0.39	0.00	1.00
Age	40706	44.56	8.86	22.00	60.00
Race	40613	0.49	0.50	0.00	1.00
Schooling	38627	13.75	0.84	8.00	14.00
Experience	40309	6.96	6.43	1.00	20.00



Tabela 4.2: Estatísticas Descritivas - Regressões Univariadas

	Coeficiente/Erro-Padrão	Constant	N	R <sup>2</sup>
Female	-3.387*** (0.482)	191.336*** (0.340)	40,964	0.001
White	16.142*** (0.500)	183.601*** (0.307)	39,751	0.026
Parental Education	7.104*** (0.098)	147.633*** (0.649)	31,103	0.143
Age	-12.267*** (0.191)	319.997*** (2.048)	41,170	0.091
Socioeconomic Index	11.285*** (0.122)	193.011*** (0.260)	27,123	0.239
Fail	-38.148*** (0.511)	200.514*** (0.270)	40,036	0.122
Preschool	32.612*** (0.605)	163.370*** (0.544)	39,432	0.069
Violência Index	-3.238*** (0.183)	188.831*** (0.250)	38,074	0.008
Quality School Index	6.482*** (0.122)	189.047*** (0.237)	39,956	0.066
Salário do Professor	0.014*** (0.000)	177.680*** (0.442)	36,953	0.030
Dummy Urbano	21.499*** (1.001)	168.695*** (0.971)	41,783	0.011
Dummy Capital	10.783*** (0.476)	183.658*** (0.332)	41,783	0.012
Federal	55.510*** (3.010)	188.550*** (0.239)	41,783	0.008
Estadual	-18.885*** (0.496)	195.354*** (0.290)	41,783	0.033
Municipal	-27.413*** (0.483)	198.593*** (0.287)	41,783	0.072
Privada	48.558*** (0.466)	174.412*** (0.255)	41,783	0.206
Legenda: *** p<0.01, ** p<0.05, * p<0.1				

Tabela 4.3: Primeiro Estágio - Regressão *Within*

Model	Within
Variável	Coeficiente/Desvio-Padrão
Socioeconomic Index	0.377 (0.233)
White	1.387** (0.665)
Female	-5.785*** (0.599)
Parental Education	0.732*** (0.152)
Age	-2.248*** (0.355)
Fail	-16.013*** (0.871)
Preschool	9.096*** (0.883)
Constante	215.027*** (4.074)
R <sup>2</sup>	0.229
N	18,438
F	135.670
$\sigma_\alpha$	32.340
$\sigma_\epsilon$	36.153
$\rho$	0.445

Legenda: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Pela inspeção destas Tabelas podemos notar que os alunos têm um desempenho médio ligeiramente superior ao das alunas. O mesmo ocorre com os brancos, em relação às demais raças. As variáveis educação dos pais, indicador socioeconômico e dummy referente à escola privada foram as que se mostraram mais capazes de explicar a variância do desempenho entre os alunos. Também é importante notar que esta tabela proporciona evidência, ainda que indireta, de que os indicadores de qualidade e violência da escola foram bem estimados, uma vez ambas as variáveis apresentaram coeficientes com o sinal esperado<sup>1</sup>.

A Tabela 4.3 mostra as estimativas do primeiro estágio do modelo. Os dados

<sup>1</sup>Ver no Apêndice a descrição do processo de estimação destas variáveis.

indicam que, controlando pelas demais características individuais, as alunas têm um desempenho médio de 5.785 pontos a menos que os alunos. De forma análoga, os alunos que declaram raça como branca obtiveram um desempenho superior que os demais em média de 1.387 pontos. Todas as variáveis apresentaram o sinal tal qual esperado, e as estimativas pontuais são semelhantes às já encontradas na literatura <sup>2</sup>.

A Tabela 4.4 reporta as estimativas dos parâmetros obtidos por meio da comparação das médias entre as diferentes das turmas. As três especificações, utilizando como controles as características do professor e do diretor, apresentaram resultados semelhantes. Basicamente, podemos afirmar que as variáveis associadas ao indivíduo, exceto a dummy de gênero e idade, apresentaram coeficientes estatisticamente diferentes de zero. Os indicadores de violência e qualidade da escola também apresentaram coeficientes significativos, bem como a maior parte das características demográficas e referentes aos professores e diretor da escola.

De posse dos parâmetros estimados nas duas regressões, podemos partir para o propósito principal desta pesquisa, isto é, a mensuração dos *peer effects*. A Tabela 4.5 abaixo reporta as estimativas dos parâmetros do modelo *linear-in-means* a partir dos coeficientes obtidos nas regressões *within* e *between groups*. Os desvios padrão foram obtidos através de *bootstrap*.

Como se pode notar, os três modelos indicam a presença de *endogenous peer-effects* <sup>3</sup>, em razão do coeficiente associado à média de desempenho dos pares ser estatisticamente diferente de zero.

Com relação aos efeitos contextuais, é interessante notar que o sinal do efeito contextual para a variável gênero tem o sinal oposto ao efeito desta variável no indivíduo. Isto é, apesar das alunas tenderem a obter um desempenho menor do que seus colegas, a proporção de alunas na sala impacta positivamente o desempenho naquela turma.

---

<sup>2</sup>Ver Pinto(2008)

<sup>3</sup>Endógeno neste contexto está associado ao fato de sua determinação ser dada como função das demais variáveis do modelo microeconômico. Assim, quando usar o termo *peer effect* endógeno, esteja claro que me refiro ao parâmetro associado ao desempenho dos pares,  $\beta_3$

Tabela 4.4: Segundo Estágio - Regressão *Between*

Variável	I	II	III
	Coeficiente/SE	Coeficiente/SE	Coeficiente/SE
Socioeconomic Index	6.235*** (0.502)	6.111*** (0.520)	6.009*** (0.555)
White	7.339*** (1.721)	7.506*** (1.779)	8.669*** (1.855)
Female	-2.293 (1.736)	-2.312 (1.778)	-3.312* (1.862)
Parental Education	0.947** (0.390)	0.793** (0.402)	0.896** (0.423)
Age	0.350 (0.723)	0.304 (0.739)	0.453 (0.771)
Fail	-19.687*** (2.223)	-20.165*** (2.267)	-20.804*** (2.347)
Preschool	15.298*** (2.192)	15.429*** (2.238)	15.679*** (2.289)
Violencia Index	-0.797** (0.354)	-0.727** (0.363)	-0.901** (0.396)
Quality School Index	0.509* (0.260)	0.606** (0.265)	0.706** (0.276)
Salário do Professor	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
Dummy Urbano	-1.107 (1.785)	-1.637 (1.832)	-1.168 (1.918)
Dummy Capital	2.114** (1.006)	2.512** (1.049)	2.978*** (1.104)
Federal	31.880*** (6.780)	19.254** (7.539)	18.144** (7.900)
Estadual	3.246*** (1.109)	3.763*** (1.167)	3.635*** (1.221)
Privada	20.330*** (1.666)	20.749*** (1.709)	18.822*** (1.821)
<i>Controls:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
Constante	158.896*** (9.083)	144.269*** (11.895)	148.121*** (14.530)
R2	0.624	0.626	0.628
N	14,400	13,517	12,266
Número de Turmas	2,761	2,590	2,355

Legenda: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Tabela 4.5: *Peer Effects* - Modelo Linear

Parâmetro	I	II	III
	Coeficiente/SE	Coeficiente/SE	Coeficiente/SE
$\beta_3$	0.405*** (0.121)	0.410*** (0.120)	0.420*** (0.120)
<i>Contextual Peer Effects</i>			
Socioeconomic Index	3.330*** (0.788)	3.225*** (0.790)	3.109*** (0.792)
White	2.976** (1.417)	3.037** (1.409)	3.642** (1.476)
Female	4.421*** (1.353)	4.422*** (1.400)	3.863*** (1.383)
Parental Education	-0.169 (0.313)	-0.264 (0.311)	-0.212 (0.323)
Age	2.456*** (0.552)	2.428*** (0.553)	2.511*** (0.568)
Fail	4.307 (2.979)	4.125 (3.056)	3.944 (3.087)
<i>Controls:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes

Legenda: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Deste modo, ao olharmos para o desempenho agregado na turma, não conseguimos verificar um efeito diferente de zero para esta variável, a proporção de alunas na turma, uma vez que efeitos de sinais opostos, o efeito contextual e o *endogenous peer effects*, estão se cancelando.

Note que o efeito marginal de  $\bar{X}_g$  no desempenho esperado da turma  $g$  é dado por <sup>4</sup>:

$$\partial \frac{E[y_{ig}]}{\partial \bar{X}_g} = \frac{\beta_1 + \beta_2}{1 - \beta_3}$$

Assim, na ausência de efeitos contextuais, o aumento na proporção de garotas deveria reduzir o desempenho de todos na turma, por meio do efeito que o desempenho menor delas causa aos demais (associado ao parâmetro  $\beta_3$ ). No entanto, o efeito contextual desta variável tem o sinal contrário e magnitude muito semelhante, de modo que o efeito marginal desta variável torna-se praticamente nulo. O mesmo ocorre com a média de idade, uma vez que embora esta variável apresente sinal negativo no desempenho individual, não pudemos encontrar evidências que turmas com média de idade mais elevada apresentassem desempenho menor.

Além disto, podemos verificar a existência de um forte componente de *peer effect* endógeno. O coeficiente  $\beta_3$  no valor de aproximadamente 0.4 implica em um *social multiplier* de 1.6 no efeito marginal de qualquer variável.

Conforme descrito na introdução, interações sociais são em geral invocadas como uma forma de se explicar o excesso de variabilidade nos *outcomes* entre os grupos que não pode ser explicada pelas diferenças nas características destes grupos. A estratégia de identificação dos parâmetros deste modelo se baseia justamente neste princípio.

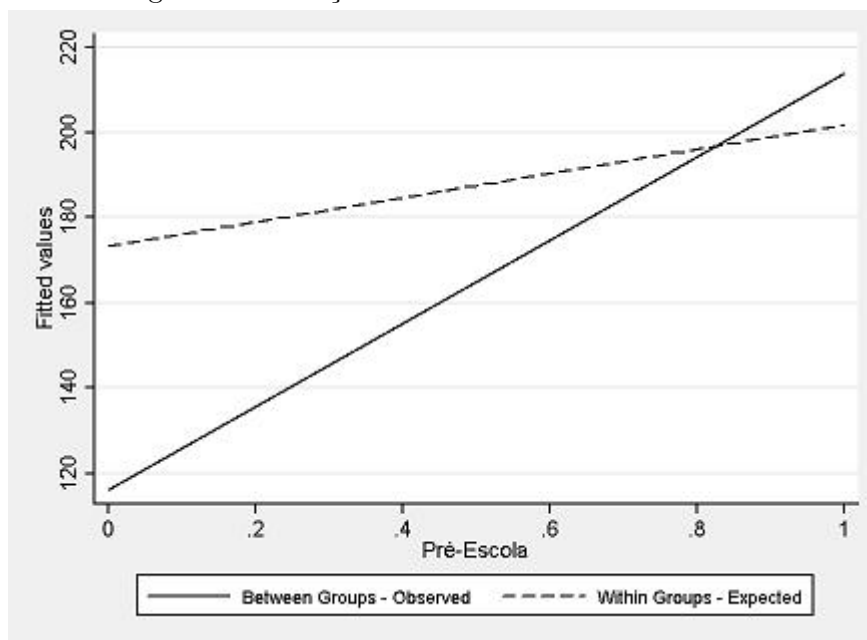
Note que através da variação *within groups* fomos capazes de encontrar o efeito marginal individual de se cursar a pré-escola. Este efeito é líquido de qualquer in-

---

<sup>4</sup>Podemos definir outros efeitos marginais de interesse, como o efeito no indivíduo  $i$  de uma alteração em  $\bar{X}_g$  que tenha sido resultado de uma alteração na característica de um indivíduo  $j \neq i$ . Neste caso, o efeito marginal será dado por  $\frac{\beta_1\beta_3 + \beta_2}{1 - \beta_3}$ .

teração social, uma vez que as médias nos grupos foram subtraídas no cálculo deste parâmetro. Na regressão entre os grupos, o efeito encontrado da variável pré-escola é significativamente superior (saltando de 9 para 15 pontos). Esta diferença entre os coeficientes, se são válidas as hipóteses assumidas, evidencia o comportamento “conformista” que caracteriza a existência de *peer effects* endógenos.

Figura 4.1: Relação entre Pré-Escola e Proficiência



Observe o gráfico 4.1. A reta contínua mostra como se comporta o desempenho esperado como função da proporção de alunos que cursou a pré-escola, tal como obtido a partir dos dados na regressão *between groups*. No entanto, nós dispomos de uma estimativa do efeito da pré-escola no desempenho do indivíduo, e a partir dela poderíamos construir uma reta que nos daria o efeito esperado de uma mudança na proporção de alunos que cursou a pré-escola, líquido de qualquer efeito de interação social. Ou seja, sobre a hipótese de que  $\beta_3$  é igual a zero, as duas retas deveriam se coincidir, uma vez que o único efeito de um aumento na proporção de alunos com pré-escola deveria ser

dado pelo efeito desta variável naquele indivíduo. A diferença entre elas é uma medida da força das interações sociais, e a partir desta diferença foi construída a estimativa de  $\beta_3$ .

Finalmente, podemos concluir a partir desta discussão que peer effects parecem ser parte importante da determinação do desempenho dos alunos na quarta série no Brasil, tanto na forma contextual – através das características dos colegas como raça, idade e indicador socioeconômico – como na forma endógena ou comportamental, por meio do desempenho dos colegas.



## 4.2 Efeitos ao Longo da Distribuição

Conforme discutido anteriormente, diversas hipóteses foram lançadas na literatura pedagógica a respeito da forma ideal de se alocar estudantes de diferentes características de forma a maximizar o desempenho do grupo. O modelo até então apresentado, embora forneça importantes *insights* sobre a magnitude e a natureza dos *peer effects*, é ainda pouco flexível para responder a este tipo de pergunta.

Sendo assim, nesta seção iremos estimar os parâmetros associados deste modelo em diversos pontos da distribuição condicional da habilidade. Deste modo, ao captar a heterogeneidade dos *peer effects* com relação à distribuição de habilidade, poderemos verificar qual hipótese melhor se adequa ao caso brasileiro, se é a “*bad apple*” ou a “*shining star*”.

Seguindo a metodologia proposta por Cooley (2007), nós iremos identificar por meio da regressão *between groups* o termo de erro do grupo,  $\xi_g$ . Plugando a estimativa de na forma reduzida do modelo, temos uma equação linear com um termo de erro simples,  $\ddot{\epsilon}_{ig}$ . Assim, condicionando a distribuição de  $\ddot{\epsilon}_{ig}$  em um determinado quantil,  $\tau$ , podemos verificar como este ponto da distribuição condicional de desempenho responde às variações nos diversos *inputs* do modelo.

$$\bar{y}_g = \frac{\beta_0}{1 - \beta_3} + \bar{X}'_g \frac{\beta_1 + \beta_2}{1 - \beta_3} + \bar{r}'_g \frac{\beta_4}{1 - \beta_3} + \kappa'_g \frac{\beta_5}{1 - \beta_3} + \frac{v_g + \bar{\epsilon}_g}{1 - \beta_3} = \pi_1 + \bar{X}'_g \pi_2 + \bar{r}'_g \pi_3 + \xi_g$$

$$\begin{aligned} y_{ig} &= \frac{\beta_0}{1 - \beta_3} + X'_{ig} \beta_1 + \bar{X}'_g \frac{\beta_3 \beta_1 + \beta_2}{1 - \beta_3} + r_{ig} \beta_4 + \bar{r}_g \frac{\beta_4 \beta_3}{1 - \beta_3} + \kappa'_g \frac{\beta_5}{1 - \beta_3} + \xi_g + \ddot{\epsilon}_{ig} = \\ &= \pi_0 + X_{ig} \pi_1 + \bar{X}'_g \pi_2 + r_{ig} \pi_3 + \bar{r}_g \pi_4 + \kappa_g \pi_5 + \hat{\xi}_g + \ddot{\epsilon}_{ig} \end{aligned}$$

As Tabelas 4.7, 4.6 e 4.8 mostram, respectivamente, as estimativas de peer effects

Tabela 4.6: Efeitos ao Longo da Distribuição  $\tau = 0.50$ 

	I	II	III
$\beta_3$	0.280** (0.138)	0.302** (0.145)	0.286* (0.165)
<i>Contextual Peer Effects</i>			
Socioeconomic Index	3.241*** (0.917)	3.055*** (0.958)	3.313*** (1.083)
White	2.844 (1.837)	2.084 (1.917)	2.104 (2.213)
Female	4.101** (1.685)	4.524** (1.807)	4.014** (2.043)
Parental Education	0.129 (0.381)	0.055 (0.407)	0.161 (0.464)
Age	1.891*** (0.649)	2.280*** (0.700)	2.585*** (0.767)
Fail	1.201 (3.639)	0.659 (4.000)	0.003 (4.556)
<i>Controls<sup>1</sup>:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
N	14,400	13,517	12,266

1: Além destes controles, todas as especificações incluem os seguintes: indicadores de qualidade, de violência, salário do professor e as dummies de urbano, capital, escola privada, estadual e municipal. Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

para os quantis 0.10, 0.50 e 0.90 da distribuição condicional de desempenho.

No que diz respeito aos efeitos contextuais, percebe-se que o efeito mediano da raça deixou de ser estatisticamente diferente de zero. As demais características do indivíduo não apresentaram grandes alterações em relação aos efeitos médios já estudados. Assim, os alunos medianos parecem responder positivamente a variações na idade da turma, na proporção de garotas <sup>5</sup>, e no indicador socioeconômico.

Com relação ao *peer effect* endógeno, nota-se que o efeito mediano foi inferior ao

<sup>5</sup>Convém, novamente, ressaltar que estas são as estimativas dos parâmetros da forma estrutural. O efeito marginal de uma alteração na proporção de alunas em uma classe contém, além do efeito contextual, o *peer effect* endógeno.

Tabela 4.7: Efeitos ao Longo da Distribuição  $\tau = 0.10$ 

	I	II	III
$\beta_3$	0.333 (0.207)	0.286 (0.244)	0.443** (0.201)
<i>Contextual Peer Effects</i>			
Socioeconomic Index	2.934** (1.273)	2.837** (1.401)	2.298* (1.208)
White	3.827 (2.526)	3.785 (2.655)	2.403 (2.454)
Female	2.713 (2.034)	3.650* (2.184)	2.726 (1.972)
Parental Education	-0.656* (0.398)	-0.529 (0.437)	-0.640 (0.395)
Age	2.923*** (0.865)	2.869*** (0.906)	3.315*** (0.879)
Fail	6.491** (3.257)	5.338 (3.810)	6.930** (3.346)
<i>Controls<sup>1</sup>:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
N	14,400	13,517	12,266

1: Além destes controles, todas as especificações incluem os seguintes: indicadores de qualidade, de violência, salário do professor e as dummies de urbano, capital, escola privada, estadual e municipal. Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Tabela 4.8: Efeitos ao Longo da Distribuição  $\tau = 0.90$ 

$\tau = 0.9$	I	II	III
$\beta_3$	0.440*** (0.127)	0.493*** (0.114)	0.493*** (0.123)
<i>Contextual Peer Effects</i>			
Socioeconomic Index	2.873*** (1.099)	2.388** (1.022)	2.144** (1.050)
White	0.237 (2.153)	0.573 (2.112)	0.369 (2.302)
Female	4.982** (2.312)	5.458** (2.170)	3.127 (2.512)
Parental Education	0.521 (0.536)	0.323 (0.483)	0.497 (0.535)
Age	1.706* (0.874)	1.761** (0.837)	2.183** (0.888)
Fail	5.802 (4.302)	7.232* (4.028)	5.687 (4.452)
<i>Controls<sup>1</sup>:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
N	14,400	13,517	12,266

1: Além destes controles, todas as especificações incluem os seguintes: indicadores de qualidade, de violência, salário do professor e as dummies de urbano, capital, escola privada, estadual e municipal. Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

efeito na média, o que sugere que a distribuição da resposta é assimétrica. Isto pode ser notado pela inspeção das Tabelas 4.7 e 4.8. Primeiramente, pode-se notar que a estimativa de *peer effect* endógeno nos quantis inferiores –  $\tau = 0.10$  – é apenas estatisticamente diferente de zero na especificação com o maior vetor de controles.

Já no quantil .9 encontramos nas 3 especificações um efeito superior aos efeitos encontrados para média, mediana e para o quantil .10. Este resultado sugere que o comportamento “conformista” é mais forte nos indivíduos no topo da distribuição de habilidade. Esta evidência corrobora a hipótese de “*one bad apple*” da literatura pedagógica, pois os alunos de maior habilidade são mais sensíveis a mudanças no comportamento do grupo do que os da base da distribuição.

### 4.3 Robustez

Nesta seção realizaremos algumas especificações alternativas com o objetivo de verificar a robustez dos resultados obtidos.

Conforme discutido na seção anterior, uma hipótese que é particularmente restritiva é a inexistência de correlação entre o indicador socioeconômico médio no grupo e os efeitos correlatos não observados,  $v_g$ . Como educação é um bem normal <sup>6</sup>, pode-se imaginar que famílias mais afortunadas comprem uma educação de melhor qualidade. Até então nós controlamos para este processo de auto-seleção na formação dos grupos por meio da hipótese de que a seleção se dá em características que nos são observáveis, que compõem o vetor de controles.

A Tabela 4.9 apresenta as estimativas de peer-effects a partir de uma ligeira modificação no modelo. Em razão da possibilidade dos nossos resultados serem guiados pela parte do processo de auto-seleção que não foi controlada pelo vetor de controles utilizado, nós instrumentalizamos o indicador socioeconômico pela regra de alocação dos alunos entre as turmas <sup>7</sup>.

A inspeção desta Tabela permite notar que esta especificação manteve os mesmos resultados do modelo anterior. As estimativas pontuais dos parâmetros mantiveram-se praticamente inalteradas, com pequenas reduções nos coeficientes associados à raça e reprovação. Além disto, o coeficiente associado ao *peer effect* endógeno, apesar da ligeira redução, deixou de ser estatisticamente diferente de zero, exceto a 10% na especificação com o maior vetor de controles.

Esta diferença na significância deve-se fortemente ao aumento no desvio-padrão das estimativas (que apresentou um aumento de mais de 100%) do que a ligeira queda na

---

<sup>6</sup>Esta afirmação é passível de questionamento. No entanto, bens medidos em níveis muito agregados como alimentação, transporte, saúde e educação comportam-se na maioria das vezes como bens normais.

<sup>7</sup> Pinto (2008) desenvolve um modelo microeconômico da decisão do diretor e demonstra as condições em que a regra de alocação dos alunos pode funcionar como instrumento para controlar a endogeneidade das características dos pares.

Tabela 4.9: *Peer Effects* - Modelo IV1

	I	II	III
$\beta_3$	0.323 (0.271)	0.362 (0.222)	0.391* (0.230)
<i>Contextual Peer Effects</i>			
Socioeconomic Index <sup>1</sup>	7.404 (8.935)	5.829 (6.988)	4.601 (6.765)
White	1.125 (3.201)	1.879 (2.572)	2.856 (2.282)
Female	4.479** (1.745)	4.395*** (1.637)	3.765** (1.656)
Parental Education	-1.407 (2.100)	-1.122 (1.645)	-0.672 (1.547)
Age	2.328*** (0.799)	2.297*** (0.729)	2.447*** (0.748)
Fail	3.282 (4.754)	3.388 (4.204)	3.574 (4.069)
<i>Controls<sup>2</sup>:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
N	13796	12934	11755

Legenda: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. 1: Instrumentalizado pela regra de alocação dos alunos nas turmas. 2: Além destes controles, todas as especificações incluem os seguintes: indicadores de qualidade, de violência, salário do professor e as dummies de urbano, capital, escola privada, estadual e municipal.

estimativa pontual do coeficiente.

Por último, é interessante notar que a estimativa do efeito do indicador socioeconômico foi substancialmente maior no modelo IV do que na estimação por OLS. Isto dá uma evidência, ainda que indireta, de que o vetor de controles utilizado controla adequadamente a auto-seleção presente no processo de formação dos grupos, uma vez que se esperava uma correlação positiva entre o indicador e o efeito correlato <sup>8</sup>.

Na Tabela 4.10 encontram-se as estimativas de *peer-effects* quando utilizamos tanto a regra de admissão quanto a regra de alocação dos alunos como instrumentos para o indicador socioeconômico. Os resultados nesta especificação reforçam as diferenças obtidas entre o modelo OLS e o modelo IV já apresentado, isto é, o coeficiente referente ao indicador socioeconômico torna-se maior, o coeficiente das demais variáveis reduz ligeiramente e os desvios padrões crescem substancialmente, o que produz estimativas pontuais não diferentes de zero para a maior parte das covariadas.

Sendo assim, iremos prosseguir as demais estimações utilizando a especificação mais simples, lembrando que esta especificação parece ser um *lower bound* no efeito contextual do indicador socioeconômico.

---

<sup>8</sup>Uma outra explicação para o aumento no coeficiente é de que o modelo IV identifica o efeito para uma subpopulação de *compliers* onde o efeito do indicador socioeconômico é maior do que na população como um todo.



Tabela 4.10: *Peer Effects* - Modelo IV2

	I	II	III
$\beta_3$	0.103 (0.438)	0.178 (0.387)	0.259 (0.332)
<i>Contextual Peer Effects</i>			
Socioeconomic Index <sup>1</sup>	17.978 (13.122)	13.923 (11.158)	10.332 (8.914)
White	-2.684 (4.062)	-0.350 (3.325)	1.496 (2.664)
Female	5.163** (2.424)	4.856** (2.287)	3.974* (2.111)
Parental Education	-4.406 (2.719)	-3.385 (2.326)	-2.215 (1.817)
Age	2.054* (1.130)	2.061* (1.076)	2.236** (1.011)
Fail	0.722 (7.161)	0.841 (6.469)	1.891 (5.343)
<i>Controls</i> <sup>2</sup> :			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
N	13,555	12,717	11,572

1: Instrumentalizado pelas regras de admissão e alocação dos alunos nas turmas. 2: Além destes controles, todas as especificações incluem os seguintes: indicadores de qualidade, de violência, salário do professor e as dummies de urbano, capital, escola privada, estadual e municipal. Legenda: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

## 4.4 Análise de Influência

Uma propriedade conhecida dos estimadores de *Ordinary Least Squares* é o fato de serem sensíveis a *outliers*, um pequeno grupo de pontos que muitas vezes é responsável por grandes alterações nas estimativas dos parâmetros. Em geral, em razão da forma de computação das estimativas de mínimos quadrados, estes pontos concentram-se nos extremos da distribuição das covariadas.

Existem várias técnicas na literatura designadas para a detecção de pontos influentes (CHATERJEE e HADI, 1986). Nesta seção, tentaremos verificar se as estimativas que encontramos na seção anterior devem-se a um fenômeno que afeta grande parte dos nossos dados ou se apenas a um pequeno grupo de observações influentes. Para isto, iremos reestimar o modelo com o vetor de controles mais completo (que inclui as características do professor e do diretor) retirando da amostra os pontos extremos da distribuição das covariadas <sup>9</sup>. Este procedimento irá nos dar indícios da robustez dos nossos resultados.

Um interessante exemplo da idéia por trás deste procedimento é por meio da introspecção sobre o que nos levou a obter tal estimativa do parâmetro referente ao *peer effect* endógeno. Conforme discutido anteriormente, na ausência de *peer effects*, devíamos observar uma queda na proficiência média da turma da magnitude do efeito no nível do indivíduo à medida que se reduzisse a proporção de alunos que cursou a pré-escola.

No entanto, imagine que as escolas onde a proporção de alunos que cursou a pré-escola é degenerada sejam essencialmente diferentes das demais. Uma escola onde nenhum aluno cursou a pré-escola provavelmente possui uma série de outras características não observáveis que reduzem o desempenho dos seus alunos (como, por exemplo, estar

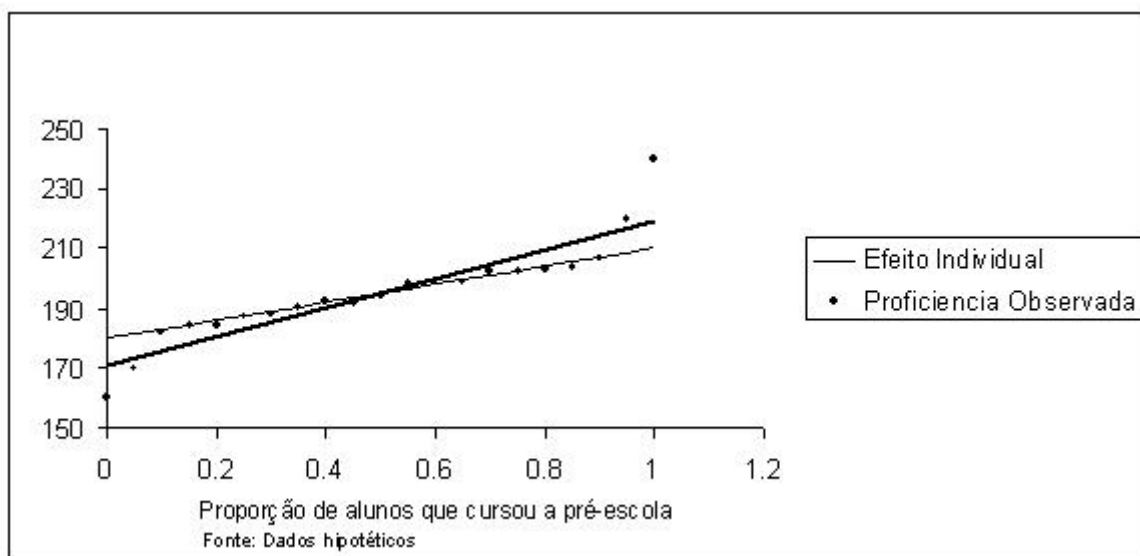
---

<sup>9</sup>Para as variáveis binárias, removemos as turmas em que a distribuição é degenerada (por exemplo, as turmas formadas apenas por mulheres). Para as variáveis que assumem mais do que dois valores foram removidas as turmas que estão acima de 95% ou abaixo de 05% na distribuição marginal da variável.

em algum lugar remoto da zona rural de um estado pouco desenvolvido). No extremo oposto, escolas em que todos os alunos cursaram a pré-escola possivelmente diferem das demais não apenas neste quesito, mas também por estarem localizadas em zonas nobres das regiões mais prósperas do país.

Deste modo, ainda que utilizemos de um vetor de controles que inclui características demográficas das escolas, é possível que estas escolas nos pontos extremos difiram das demais não apenas em observáveis, mas também em não observáveis. A título de ilustração, observe a figura 4.2. Note que a proficiência observada nas turmas é consistente com a hipótese de inexistência de *peer effects*, uma vez que a proficiência média observada nas turmas cresce exatamente na magnitude que seria implicada pelo crescimento da proficiência do aluno que cursou a pré-escola.

Figura 4.2: Estimação de *peer effects* na presença de *outliers*



No entanto, em razão dos *outliers*, a estimativa - representada no gráfico pela linha mais escura - por meio da regressão *between groups* do efeito de um acréscimo na proporção de alunos que cursou a pré-escola será maior do que o efeito individual,

Tabela 4.11: Análise de Influência - Modelo Linear

	0	I	II	III	IV	V	VI	VII
$\beta_3$	0.420*** (0.120)	0.418*** (0.055)	0.458*** (0.052)	0.414*** (0.056)	0.454*** (0.052)	0.409*** (0.056)	0.445*** (0.053)	0.426*** (0.055)
<i>Contextual Peer Effects</i>								
Socioeconomic Index	3.109*** (0.792)	1.983*** (0.364)	2.745*** (0.419)	2.695*** (0.415)	2.921*** (0.432)	3.313*** (0.463)	2.644*** (0.411)	3.074*** (0.444)
White	3.642*** (1.476)	3.413*** (0.923)	3.635*** (0.936)	3.825*** (0.948)	2.307*** (0.862)	3.193*** (0.910)	3.486*** (0.927)	3.586*** (0.923)
Female	3.863*** (1.383)	4.090*** (0.588)	3.775*** (0.598)	3.942*** (0.593)	3.743*** (0.599)	3.815*** (0.597)	3.785*** (0.598)	3.664*** (0.602)
Parental Education	-0.212 (0.323)	0.015 (0.183)	-0.162 (0.176)	-0.136 (0.177)	-0.257 (0.173)	-0.421** (0.169)	-0.156 (0.176)	-0.237 (0.174)
Age	2.511*** (0.568)	2.094*** (0.320)	2.434*** (0.320)	2.278*** (0.320)	2.130*** (0.320)	2.292*** (0.320)	2.050*** (0.320)	1.734*** (0.322)
Fail	3.944*** (3.087)	7.003*** (1.283)	5.274*** (1.405)	5.648*** (1.378)	5.042*** (1.422)	3.426** (1.544)	5.339*** (1.400)	5.169*** (1.413)

Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . 0: *Baseline*. I-VII: Regressões sem os extremos da distribuição marginal de pré-escola, indicador socioeconômico, branco, dummy de gênero, educação dos pais, idade e reprovação; respectivamente. Todas as especificações incluem os seguintes controles: indicadores de qualidade, de violência, salário do professor, as dummies de urbano, capital, escola privada, estadual, municipal, além das características do professor e diretor.

indicando erroneamente a presença de peer effects.

Deste modo, ao retirar os pontos extremos, estaremos nos certificando que os parâmetros estimados referem-se a efeitos que ocorrem efetivamente por meio das mudanças na distribuição de desempenho causadas por mudanças suaves, e não nos pontos extremos, das covariadas. A Tabela 4.11 reporta os resultados obtidos:

Como se pode notar, todas as estimativas de peer effects, tanto contextual quanto endógeno, se mostraram robustas a exclusão dos pontos extremos da distribuição das covariadas.

Realizando o mesmo procedimento para os efeitos nos quantis condicionais de desempenho, temos nas Tabelas 4.13, 4.12 e 4.14 os parâmetros estimados para o quantil .10, a mediana e para o quantil .90:

Como se pode notar, as estimativas dos parâmetros se mantiveram praticamente as mesmas para os quantis .90 e .10. Por outro lado, não verificamos robustez do efeito da proporção de alunas no modelo para o quantil .50, sendo que em quase todas as sub-

Tabela 4.12: Análise de Influência -  $\tau = 0.50$ 

	0	I	II	III	IV	V	VI	VII
$\beta_3$	0.286* (0.165)	0.231 (0.178)	0.264 (0.162)	0.313** (0.156)	0.280* (0.160)	0.275 (0.169)	0.299* (0.158)	0.316* (0.168)
<i>Contextual Peer Effects</i>								
Socioeconomic Index	3.313*** (1.083)	3.799*** (1.182)	3.251*** (1.040)	3.010*** (1.009)	2.918*** (0.999)	3.334*** (1.087)	2.376*** (0.919)	1.641* (0.920)
White	2.104 (2.213)	2.595 (2.279)	1.290 (2.109)	2.271 (2.237)	2.313 (2.230)	2.828 (2.340)	2.516 (2.282)	1.904 (2.614)
Female	4.014** (2.043)	3.951* (2.102)	3.250 (2.094)	3.499 (2.144)	2.695 (2.144)	3.146 (2.148)	3.237 (2.149)	2.613 (2.517)
Parental Education	0.161 (0.464)	0.009 (0.451)	0.199 (0.464)	0.225 (0.474)	0.246 (0.472)	0.227 (0.483)	0.435 (0.489)	0.371 (0.546)
Age	2.585*** (0.767)	2.535*** (0.781)	2.311*** (0.794)	2.479*** (0.781)	2.006** (0.814)	1.458 (0.935)	2.282*** (0.762)	1.789** (0.848)
Fail	0.003 (4.556)	-1.247 (4.867)	0.304 (4.336)	1.233 (4.257)	0.102 (4.321)	1.837 (4.324)	2.520 (4.003)	1.257 (4.139)
N	12,266	11,956	11,507	12,172	11,693	12,135	10,251	8,534

Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . 0: *Baseline*. I-VII: Regressões sem os extremos da distribuição marginal de pré-escola, indicador socioeconômico, branco, dummy de gênero, educação dos pais, idade e reprovação; respectivamente. Todas as especificações incluem os seguintes controles: indicadores de qualidade, de violência, salário do professor, as dummies de urbano, capital, escola privada, estadual, municipal, além das características do professor e diretor.

Tabela 4.13: Análise de Influência -  $\tau = 0.10$ 

	0	I	II	III	IV	V	VI	VII
$\beta_3$	0.443** (0.201)	0.427** (0.198)	0.484** (0.196)	0.445** (0.190)	0.440** (0.196)	0.434** (0.173)	0.512*** (0.161)	0.450** (0.197)
<i>Contextual Peer Effects</i>								
Socioeconomic Index	2.298* (1.208)	2.409** (1.185)	2.102* (1.215)	2.305** (1.106)	2.133* (1.102)	2.456** (1.078)	2.128** (1.013)	0.987 (0.906)
White	2.403 (2.454)	2.297 (2.350)	1.652 (2.388)	2.929 (2.452)	1.576 (2.290)	2.345 (2.149)	1.122 (2.097)	2.285** (0.997)
Female	2.726 (1.972)	2.022 (1.927)	3.295 (2.020)	3.108 (2.016)	2.161 (2.042)	3.389* (1.810)	1.536 (1.936)	4.184** (1.677)
Parental Education	-0.640 (0.395)	-0.878** (0.377)	-0.881** (0.405)	-0.635 (0.391)	-0.641 (0.405)	-0.793** (0.353)	-0.725* (0.372)	-0.511 (0.424)
Age	3.315*** (0.879)	3.119*** (0.869)	2.925*** (0.957)	3.173*** (0.876)	3.042*** (0.917)	2.976*** (0.901)	2.994*** (0.856)	2.978*** (0.842)
Fail	6.930** (3.346)	6.265* (3.388)	7.055** (3.438)	7.391** (3.133)	5.978* (3.443)	6.346** (3.009)	8.272*** (2.776)	7.524** (3.129)
N	12,266	11,956	11,507	12,172	11,693	12,135	10,251	8,534

Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . 0: *Baseline*. I-VII: Regressões sem os extremos da distribuição marginal de pré-escola, indicador socioeconômico, branco, dummy de gênero, educação dos pais, idade e reprovação; respectivamente. Todas as especificações incluem os seguintes controles: indicadores de qualidade, de violência, salário do professor, as dummies de urbano, capital, escola privada, estadual, municipal, além das características do professor e diretor.

Tabela 4.14: Análise de Influência -  $\tau = 0.90$ 

	0	I	II	III	IV	V	VI	VII
$\beta_3$	0.493*** (0.123)	0.506*** (0.118)	0.494*** (0.111)	0.507*** (0.114)	0.502*** (0.113)	0.515*** (0.101)	0.479*** (0.121)	0.438*** (0.144)
<i>Contextual Peer Effects</i>								
Socioeconomic Index	2.144** (1.050)	2.101** (1.041)	2.150** (0.991)	1.997** (0.994)	1.894** (0.945)	1.984** (0.880)	1.912* (1.009)	1.967* (1.116)
White	0.369 (2.302)	0.659 (2.303)	-1.133 (2.160)	0.355 (2.273)	0.800 (2.327)	0.094 (2.008)	1.378 (2.434)	4.087*** (1.328)
Female	3.127 (2.512)	4.241* (2.367)	2.277 (2.482)	1.989 (2.677)	2.697 (2.505)	2.584 (2.248)	1.445 (2.609)	7.734*** (2.370)
Parental Education	0.497 (0.535)	0.298 (0.498)	0.337 (0.508)	0.314 (0.516)	0.426 (0.518)	0.388 (0.467)	0.411 (0.516)	0.398 (0.629)
Age	2.183** (0.888)	2.016** (0.882)	1.972** (0.858)	2.186** (0.884)	1.795** (0.886)	1.725** (0.855)	2.416*** (0.860)	1.742* (0.940)
Fail	5.687 (4.452)	7.252* (4.270)	7.678* (3.928)	6.052 (4.244)	6.262 (4.131)	7.565** (3.551)	5.263 (4.335)	12.551*** (3.934)
N	12,266	11,956	11,507	12,172	11,693	12,135	10,251	8,534

Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . 0: *Baseline*. I-VII: Regressões sem os extremos da distribuição marginal de pré-escola, indicador socioeconômico, branco, dummy de gênero, educação dos pais, idade e reprovação; respectivamente. Todas as especificações incluem os seguintes controles: indicadores de qualidade, de violência, salário do professor, as dummies de urbano, capital, escola privada, estadual, municipal, além das características do professor e diretor.

amostras o efeito estimado foi não estatisticamente diferente de zero. Ainda assim, de um modo geral, os efeitos estimados nas sub-amostras foram muito semelhantes àqueles encontrados na amostra completa.

## 4.5 Extensões

Uma limitação do modelo *linear-in-means* é o fato que ele não permite a flexibilidade necessária para que seja possível pensar em realocações maximizadoras de desempenho. Isto ocorre porque, em razão do modelo ser linear, os ganhos obtidos por um grupo através de uma política de realocação serão necessariamente compensados pelas perdas dos demais. Deste modo, embora o modelo *linear-in-means* seja capaz de evidenciar importantes elementos da relação entre o desempenho de um aluno e o grupo no qual ele está inserido, é necessário que se flexibilize ligeiramente a sua forma funcional para que seja possível responder a algumas perguntas.

Deste modo, nesta seção flexibilizaremos o modelo *linear-in-means*, permitindo que o desempenho de um aluno seja relacionado com as características dos seus colegas por meio de um polinômio de segunda ordem. Nós também permitimos que a variância de  $X$  e  $y$  interfiram no desempenho dos demais. Assim, permitimos que o desempenho de um aluno seja influenciado de forma não linear pela sua idade, de forma não linear pela idade média dos seus colegas, pela variância da idade dos seus colegas, além da média e variância de  $y$  naquela turma. Matematicamente <sup>10</sup>:

$$y_{ig} = \beta_0 + X'_{ig}\beta_1 + \bar{X}'_g\beta_2 + \beta_3\bar{y}_g + r_{ig}\beta_4 + X'^2_{ig}\beta_5 + s^2_{X_g}\beta_6 + \bar{X}'^2_g\beta_7 + s^2_{y_g}\beta_8 + \kappa'_g\beta_9 + v_g + \epsilon_{ig}$$

Onde  $s^2_{X_g}$  e  $s^2_{y_g}$  são o estimadores de máxima verossimilhança da variância de  $X$  e  $y$  dentro da turma. A Tabela 4.15 sumariza os resultados encontrados.

Note que neste modelo, as variáveis binárias aparecem apenas com o termo linear

---

<sup>10</sup>Note que não é necessário incluir todos estes termos para as variáveis binárias.

Tabela 4.15: *Peer Effects* - Modelo Estendido

	I	II	III
$\beta_3$	0.515*** (0.087)	0.508*** (0.093)	0.507*** (0.098)
<i>Contextual Peer Effects</i>			
<i>Linear</i>			
Socioeconomic Index	2.061*** (0.705)	2.147*** (0.751)	2.178*** (0.774)
White	-1.372 (2.817)	-0.945 (2.949)	-1.247 (3.077)
Female	-7.010 (5.789)	-8.398 (6.192)	-7.624 (6.400)
Parental Education	-4.262*** (1.396)	-3.958*** (1.440)	-3.872*** (1.488)
Age	16.965*** (6.333)	17.799*** (6.581)	18.657*** (6.857)
Fail	-6.887 (5.323)	-5.516 (5.395)	-5.883 (5.708)
<i>Squared</i>			
Socioeconomic Index	0.278** (0.109)	0.272** (0.113)	0.287** (0.119)
White	1.735 (3.002)	1.522 (3.137)	2.031 (3.281)
Female	12.989** (5.689)	14.688** (6.079)	13.654** (6.308)
Parental Education	0.367*** (0.125)	0.346*** (0.129)	0.348*** (0.134)
Age	-0.613** (0.271)	-0.636** (0.281)	-0.676** (0.293)
Fail	14.510*** (4.546)	13.101*** (4.577)	13.736*** (4.878)
<i>Variance</i>			
Achievement	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
Socioeconomic Index	-0.404 (0.249)	-0.442* (0.264)	-0.429 (0.277)
Age	-1.198*** (0.398)	-1.279*** (0.422)	-1.275*** (0.434)
Parental Education	-0.033 (0.138)	-0.067 (0.145)	-0.053 (0.150)
<i>Controls<sup>1</sup>:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
N	14,400	13,517	12,266

1: Além destes controles, todas as especificações incluem os seguintes: indicadores de qualidade, de violência, salário do professor e as dummies de urbano, capital, escola privada, estadual e municipal.

Legenda: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.



e quadrático referente à proporção dentro do grupo. Assim o efeito marginal no desempenho esperado na turma de uma mudança em  $X$  em uma destas variáveis é dado por:

$$\partial \frac{E[\bar{y}_g]}{\partial \bar{X}_g} = \frac{\beta_1 + \beta_2 + 2\beta_7 \bar{X}_g}{1 - \beta_3}$$

Deste modo, temos uma perspectiva muito mais rica sobre os efeitos de alterações na composição dos pares do que aquela obtida através do modelo puramente linear. A inspeção da Tabela permite observar diversos resultados interessantes. Primeiramente, percebe-se que exceto pela variável idade, todas as demais possuem efeitos de segunda ordem de sinal positivo. Isto significa que, sobre algumas condições, as políticas de segregação por tipos pode ser ótima <sup>11</sup>. Além disto, podemos verificar que a heterogeneidade etária dentro da turma apresentou um efeito negativo sobre o desempenho dos alunos. Além disto, podemos verificar um pequeno, mas estatisticamente diferente de zero, efeito da dispersão de proficiência na turma sobre o desempenho. Deste modo, obtivemos evidências de que a hipótese conhecida como *rainbow* é a mais adequada para o caso brasileiro. Também é importante ressaltar que esta especificação a magnitude do *peer effect* endógeno subiu significativamente, implicando em um *social multiplier* da magnitude de 2.

Já os efeitos nos diferentes pontos da distribuição condicional podem ser verificados nas Tabelas 4.16, 4.17 e 4.18:

---

<sup>11</sup>A definição da forma ótima de alocação dos alunos dependerá, também, da distribuição conjunta de  $X$ . Sobre este tópico, ver Arnott e Rowse (1987) e Graham, Imbens e Ridder (2008)

Tabela 4.16: Modelo Estendido -  $\tau = 0.10$ 

$\tau = 0.10$	I	II	III
$\beta_3$	0.556*** (0.096)	0.536*** (0.127)	0.618*** (0.099)
<i>Contextual Peer Effects</i>			
<i>Linear</i>			
Socioeconomic Index	1.688** (0.668)	1.606** (0.784)	1.082* (0.651)
White	1.774 (2.400)	3.433 (3.103)	3.444 (2.647)
Female	-6.109 (4.835)	-5.265 (5.884)	-5.393 (4.759)
Parental Education	-3.069** (1.295)	-2.647* (1.541)	-2.275* (1.328)
Age	12.271** (5.742)	12.270* (6.902)	12.104** (6.064)
Fail	-0.051 (3.509)	-0.671 (4.532)	0.792 (3.764)
<i>Squared</i>			
Socioeconomic Index	0.166 (0.107)	0.165 (0.130)	0.208* (0.120)
White	1.141 (2.401)	-0.197 (2.929)	-0.469 (2.332)
Female	10.162** (4.823)	9.665 (5.878)	9.217* (4.804)
Parental Education	0.267** (0.117)	0.244* (0.141)	0.202* (0.122)
Age	-0.401 (0.246)	-0.395 (0.296)	-0.387 (0.260)
Fail	4.297 (3.072)	4.940 (3.823)	3.924 (3.041)
<i>Variance</i>			
Achievement	-0.004*** (0.001)	-0.004*** (0.001)	-0.003*** (0.001)
Socioeconomic Index	-0.330 (0.206)	-0.253 (0.243)	-0.163 (0.201)
Age	0.131 (0.333)	0.137 (0.402)	0.105 (0.338)
Parental Education	0.208* (0.121)	0.166 (0.144)	0.123 (0.124)
<i>Controls<sup>1</sup>:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
N	14,400	13,517	12,266

1: Além destes controles, todas as especificações incluem os seguintes: indicadores de qualidade, de violência, salário do professor e as dummies de urbano, capital, escola privada, estadual e municipal.

Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Tabela 4.17: Modelo Estendido -  $\tau = 0.50$ 

$\tau.50$	I	II	III
$\beta_3$	0.186 (0.154)	0.229 (0.160)	0.247 (0.170)
<i>Contextual Peer Effects</i>			
<i>Linear</i>			
Socioeconomic Index	2.405*** (0.893)	2.124** (0.907)	2.060** (0.952)
White	2.456 (3.749)	1.854 (3.971)	2.489 (4.313)
Female	2.257 (7.218)	3.522 (7.590)	4.791 (8.106)
Parental Education	-3.676** (1.716)	-3.305* (1.818)	-3.730* (1.951)
Age	10.641 (7.512)	12.842 (8.020)	12.817 (8.581)
Fail	0.658 (5.290)	2.293 (5.435)	2.337 (5.839)
<i>Squared</i>			
Socioeconomic Index	0.783*** (0.142)	0.738*** (0.151)	0.775*** (0.164)
White	1.412 (3.969)	1.722 (4.207)	0.270 (4.453)
Female	7.171 (7.127)	6.464 (7.471)	4.454 (7.973)
Parental Education	0.428*** (0.158)	0.395** (0.166)	0.444** (0.182)
Age	-0.365 (0.321)	-0.444 (0.343)	-0.445 (0.367)
Fail	-1.833 (4.858)	-2.691 (5.095)	-2.388 (5.428)
<i>Variance</i>			
Achievement	0.006*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
Socioeconomic Index	-0.291 (0.314)	-0.188 (0.328)	-0.159 (0.353)
Age	-0.335 (0.457)	-0.476 (0.487)	-0.605 (0.519)
Parental Education	0.159 (0.165)	0.082 (0.177)	0.118 (0.188)
<i>Controls<sup>1</sup>:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
N	14,400	13,517	12,266

1: Além destes controles, todas as especificações incluem os seguintes: indicadores de qualidade, de violência, salário do professor e as dummies de urbano, capital, escola privada, estadual e municipal.

Legenda: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Tabela 4.18: Modelo Estendido -  $\tau = 0.90$ 

$\tau = 0.90$	I	II	III
$\beta_3$	0.306 (0.215)	0.373** (0.153)	0.375** (0.164)
<i>Contextual Peer Effects</i>			
<i>Linear</i>			
Socioeconomic Index	1.984 (1.297)	1.562 (0.964)	1.346 (0.985)
White	4.172 (5.001)	0.977 (4.071)	2.487 (4.472)
Female	18.698* (9.636)	14.055* (7.689)	15.244* (8.492)
Parental Education	-1.983 (2.209)	-1.913 (1.873)	-1.925 (2.067)
Age	5.450 (9.832)	10.614 (8.479)	9.216 (8.948)
Fail	7.736 (6.188)	9.066* (4.964)	6.422 (5.728)
<i>Squared</i>			
Socioeconomic Index	0.742*** (0.205)	0.685*** (0.169)	0.732*** (0.183)
White	-4.420 (5.061)	-0.785 (4.225)	-2.082 (4.528)
Female	-7.514 (9.066)	-3.740 (7.355)	-5.986 (8.223)
Parental Education	0.315 (0.194)	0.272 (0.167)	0.289 (0.182)
Age	-0.167 (0.419)	-0.382 (0.361)	-0.326 (0.380)
Fail	-2.480 (6.194)	-2.711 (5.168)	-0.109 (5.506)
<i>Variance</i>			
Achievement	0.017*** (0.005)	0.016*** (0.004)	0.015*** (0.004)
Socioeconomic Index	0.226 (0.355)	0.190 (0.304)	0.217 (0.328)
Age	-0.458 (0.580)	-0.766 (0.504)	-0.598 (0.532)
Parental Education	0.128 (0.214)	0.073 (0.184)	0.095 (0.199)
<i>Controls<sup>1</sup>:</i>			
<i>Dummy Estado</i>	Yes	Yes	Yes
<i>Professor Characteristics</i>	No	Yes	Yes
<i>Principal Characteristics</i>	No	No	Yes
N	14,400	13,517	12,266

1: Além destes controles, todas as especificações incluem os seguintes: indicadores de qualidade, de violência, salário do professor e as dummies de urbano, capital, escola privada, estadual e municipal.

Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Podemos notar, para os efeitos medianos, que apenas as variáveis indicador socioeconômico e educação dos pais apresentaram efeitos de segundo grau estatisticamente diferentes de zero. Tal qual o efeito na média, podemos verificar um pequeno efeito positivo da dispersão de proficiência no desempenho dos alunos, o que provê mais evidência a favor da hipótese de *rainbow*. Com relação à dispersão das demais variáveis - educação dos pais, indicador socioeconômico e idade - não foi possível encontrar nenhum efeito significativo.

Para o quantil 0.10, verificamos que a dummy de gênero, o indicador socioeconômico e a educação dos pais foram estatisticamente diferentes de zero. Aqui foi possível notar uma heterogeneidade interessante, pois a dispersão de desempenho mostrou-se de efeito negativo. Isto indica que, embora em média os alunos ganhem ao estarem em turmas mais heterogêneas, estes ganhos não são distribuídos de forma homogênea na turma, havendo inclusive perdas para alguns. Este tipo de evidência torna claro o *trade-off* defrontado pelos *policy makers* ao desenhar a composição dos grupos. Também é interessante notar que nesta especificação a magnitude dos *peer effects* endógenos foi bem elevada, mesmo para os alunos na base da distribuição de desempenho.

Para os alunos no topo da distribuição de habilidade, verificamos que apenas o indicador socioeconômico apresentou um termo de segunda ordem estatisticamente diferente de zero. A dispersão de desempenho dentro da turma apresentou sinal positivo, o que significa que estes alunos são beneficiados por turmas mais heterogêneas.

Novamente, buscando verificar se os resultados encontrados devem-se a presença de alguns *outliers*, reestimamos os modelos retirando da amostra os extremos da distribuição das covariadas. As Tabelas 4.19, 4.20, 4.21 e 4.22 sumarizam os resultados obtidos para o efeito médio, mediano e nos quantis 0.10 e 0.90:

Podemos notar que para os efeitos médios, os resultados obtidos nas sub-amostras foram essencialmente os mesmos encontrados na amostra completa. Para o quantil .10, verificamos que o efeito não linear referente à educação dos pais não se mostrou

robusto a estimação nas sub-amostras. O mesmo ocorreu com o efeito quadrático da proporção de mulheres no quantil .90, cuja significância desapareceu em várias das especificações. Já os demais resultados permaneceram essencialmente inalterados nas regressões realizadas nas sub-amostras.

Tabela 4.19: Análise de Influência - Modelo Estendido

	0	I	II	III	IV	V	VI	VII
$\beta_3$	0.507*** (0.098)	0.539*** (0.091)	0.501*** (0.105)	0.518*** (0.095)	0.522*** (0.093)	0.502*** (0.099)	0.503*** (0.108)	0.488*** (0.115)
<i>Contextual Peer Effects</i>								
<i>Linear</i>								
Socioeconomic Index	2.178*** (0.774)	2.075*** (0.751)	1.998** (0.808)	1.993*** (0.739)	2.006*** (0.732)	2.230*** (0.784)	2.157*** (0.812)	1.670** (0.813)
White	-1.247 (3.077)	-1.423 (2.931)	-3.104 (4.165)	-0.722 (3.052)	-1.342 (3.034)	-1.583 (3.131)	-0.719 (3.428)	-3.091 (3.787)
Female	-7.624 (6.400)	-8.238 (6.166)	-5.983 (6.922)	-2.082 (7.137)	-7.890 (6.412)	-7.428 (6.556)	-5.349 (7.206)	-4.784 (8.312)
Parental Education	-3.872*** (1.488)	-4.206*** (1.460)	-4.126*** (1.587)	-3.818** (1.483)	-3.719** (1.535)	-3.675** (1.531)	-2.576 (1.597)	-3.227* (1.770)
Age	18.657*** (6.857)	17.332*** (6.623)	18.909*** (7.239)	19.440*** (6.873)	19.633*** (7.144)	30.435*** (10.921)	17.792** (7.469)	13.652* (7.732)
Fail	-5.883 (5.708)	-4.130 (5.290)	-4.959 (5.943)	-5.198 (5.536)	-4.287 (5.319)	-7.971 (6.151)	-6.502 (6.859)	-1.125 (5.619)
<i>Squared</i>								
Socioeconomic Index	0.287** (0.119)	0.256** (0.125)	0.317** (0.125)	0.286** (0.119)	0.215* (0.124)	0.285** (0.121)	0.270* (0.139)	0.216 (0.161)
White	2.031 (3.281)	1.592 (3.111)	3.044 (4.569)	1.623 (3.243)	2.351 (3.285)	2.677 (3.366)	1.798 (3.941)	2.753 (4.368)
Female	13.654** (6.308)	14.323** (6.096)	11.748* (6.745)	8.270 (6.996)	14.040** (6.355)	13.114** (6.436)	12.803* (7.249)	11.351 (8.227)
Parental Education	0.348*** (0.134)	0.364*** (0.131)	0.382*** (0.144)	0.345*** (0.134)	0.335** (0.138)	0.328** (0.137)	0.243* (0.143)	0.309* (0.165)
Age	-0.676** (0.293)	-0.629** (0.284)	-0.691** (0.310)	-0.708** (0.294)	-0.721** (0.307)	-1.226** (0.490)	-0.637** (0.318)	-0.473 (0.327)
Fail	13.736*** (4.878)	12.572*** (4.541)	13.371*** (5.183)	13.625*** (4.816)	12.346*** (4.699)	17.354*** (5.649)	16.360** (6.385)	8.321 (5.132)
<i>Variance</i>								
Achievement	0.003*** (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.005*** (0.001)
Socioeconomic Index	-0.429 (0.277)	-0.382 (0.265)	-0.391 (0.289)	-0.373 (0.269)	-0.375 (0.266)	-0.416 (0.279)	-0.398 (0.301)	-0.261 (0.314)
Age	-1.275*** (0.434)	-1.009** (0.397)	-1.218*** (0.450)	-1.231*** (0.426)	-1.228*** (0.424)	-1.180*** (0.434)	-1.081** (0.435)	-0.758* (0.440)
Parental Education	-0.053 (0.150)	-0.038 (0.144)	-0.033 (0.157)	-0.051 (0.148)	-0.049 (0.147)	-0.072 (0.154)	-0.115 (0.164)	-0.076 (0.171)
N	12,266	11,956	11,507	12,172	11,693	12,135	10,251	8,534

Legenda: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. 0: *Baseline*. I-VII: Regressões sem os extremos da distribuição marginal de pré-escola, indicador socioeconômico, branco, dummy de gênero, educação dos pais, idade e reprovação; respectivamente. Todas as especificações incluem os seguintes controles: indicadores de qualidade, de violência, salário do professor, as dummies de urbano, capital, escola privada, estadual, municipal, além das características do professor e diretor.

Tabela 4.20: Análise de Influência - Modelo Estendido -  $\tau = 0.10$ 

$\tau = 0.10$	0	I	II	III	IV	V	VI	VII
$\beta_3$	0.618*** (0.099)	0.604*** (0.110)	0.652*** (0.092)	0.657*** (0.086)	0.588*** (0.112)	0.600*** (0.098)	0.589*** (0.120)	0.598*** (0.131)
<i>Contextual Peer Effects</i>								
<i>Linear</i>								
Socioeconomic Index	1.082* (0.651)	1.299* (0.717)	0.968 (0.602)	0.873 (0.577)	1.159* (0.670)	1.243** (0.622)	1.265* (0.698)	1.094 (0.794)
White	3.444 (2.647)	3.144 (2.784)	2.840 (2.927)	3.783 (2.469)	3.646 (2.885)	3.687 (2.620)	3.083 (3.124)	3.872 (3.650)
Female	-5.393 (4.759)	-5.109 (4.986)	-5.070 (4.762)	-3.179 (4.784)	-6.222 (5.470)	-6.549 (4.975)	-6.330 (5.779)	-2.223 (6.872)
Parental Education	-2.275* (1.328)	-2.180 (1.372)	-2.061 (1.335)	-2.067* (1.219)	-2.522* (1.448)	-2.479* (1.314)	-1.850 (1.476)	-0.445 (1.737)
Age	12.104** (6.064)	11.733* (6.168)	14.222** (5.967)	12.825** (5.778)	14.741** (6.756)	21.360** (8.496)	17.754** (7.104)	9.637 (7.744)
Fail	0.792 (3.764)	0.867 (3.989)	3.196 (3.338)	2.093 (3.249)	1.704 (3.893)	0.800 (3.691)	1.005 (4.636)	0.390 (4.728)
<i>Squared</i>								
Socioeconomic Index	0.208* (0.120)	0.262** (0.129)	0.234* (0.123)	0.205* (0.114)	0.225* (0.131)	0.256** (0.118)	0.240* (0.142)	0.192 (0.180)
White	-0.469 (2.332)	-1.378 (2.472)	-1.096 (2.808)	-1.198 (2.081)	-0.067 (2.614)	-0.282 (2.326)	-1.009 (3.057)	-2.473 (3.469)
Female	9.217* (4.804)	8.919* (5.081)	8.744* (4.700)	7.010 (4.748)	9.911* (5.503)	10.055** (4.990)	10.496* (5.954)	7.194 (6.856)
Parental Education	0.202* (0.122)	0.193 (0.125)	0.175 (0.122)	0.173 (0.111)	0.238* (0.133)	0.213* (0.120)	0.186 (0.135)	0.048 (0.158)
Age	-0.387 (0.260)	-0.385 (0.265)	-0.487* (0.256)	-0.419* (0.249)	-0.506* (0.292)	-0.832** (0.382)	-0.628** (0.303)	-0.266 (0.328)
Fail	3.924 (3.041)	3.594 (3.159)	2.871 (2.793)	3.357 (2.639)	2.772 (3.310)	4.488 (3.281)	4.221 (4.245)	5.337 (4.273)
<i>Variance</i>								
Achievement	-0.003*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)	-0.002*** (0.001)
Socioeconomic Index	-0.163 (0.201)	-0.133 (0.208)	-0.105 (0.188)	-0.131 (0.180)	-0.034 (0.210)	-0.107 (0.194)	-0.145 (0.244)	0.013 (0.252)
Age	0.105 (0.338)	0.080 (0.354)	-0.030 (0.338)	0.030 (0.303)	-0.053 (0.354)	-0.124 (0.325)	-0.224 (0.360)	0.036 (0.411)
Parental Education	0.123 (0.124)	0.124 (0.127)	0.090 (0.121)	0.109 (0.113)	0.170 (0.130)	0.145 (0.121)	0.090 (0.136)	0.012 (0.156)
N	12,266	11,956	11,507	12,172	11,693	12,135	10,251	8,534

Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . 0: *Baseline*. I-VII: Regressões sem os extremos da distribuição marginal de pré-escola, indicador socioeconômico, branco, dummy de gênero, educação dos pais, idade e reprovação; respectivamente. Todas as especificações incluem os seguintes controles: indicadores de qualidade, de violência, salário do professor, as dummies de urbano, capital, escola privada, estadual, municipal, além das características do professor e diretor.



Tabela 4.21: Análise de Influência - Modelo Estendido -  $\tau = 0.50$ 

$\tau = 0.50$	0	I	II	III	IV	V	VI	VII
$\beta_3$	0.247 (0.170)	0.219 (0.172)	0.272* (0.160)	0.299* (0.154)	0.212 (0.189)	0.232 (0.152)	0.184 (0.210)	0.326*** (0.121)
<i>Contextual Peer Effects</i>								
<i>Linear</i>								
Socioeconomic Index	2.060** (0.952)	2.361** (0.994)	1.583* (0.866)	1.597* (0.844)	2.065** (1.005)	2.193*** (0.845)	2.186** (1.053)	0.910 (0.579)
White	2.489 (4.313)	2.763 (4.282)	3.512 (5.373)	3.536 (4.223)	2.583 (4.785)	2.016 (3.847)	3.218 (5.085)	3.387 (3.491)
Female	4.791 (8.106)	1.389 (8.129)	3.400 (8.394)	5.985 (9.320)	3.642 (9.292)	5.752 (7.378)	5.396 (9.945)	12.457* (7.337)
Parental Education	-3.730* (1.951)	-3.796** (1.925)	-4.130** (1.955)	-3.481* (1.901)	-3.126 (2.225)	-2.765 (1.753)	-2.578 (2.176)	-2.445 (1.530)
Age	12.817 (8.581)	7.638 (8.431)	7.579 (8.663)	15.236* (8.484)	19.501* (10.162)	25.379** (12.440)	14.503 (9.801)	8.211 (6.706)
Fail	2.337 (5.839)	3.009 (5.679)	4.977 (5.463)	3.748 (5.462)	1.385 (6.472)	1.798 (5.388)	7.883 (7.089)	5.537 (4.124)
<i>Squared</i>								
Socioeconomic Index	0.775*** (0.164)	0.783*** (0.171)	0.793*** (0.167)	0.737*** (0.161)	0.626*** (0.175)	0.769*** (0.148)	0.589*** (0.184)	0.507*** (0.139)
White	0.270 (4.453)	-0.793 (4.449)	-4.216 (5.717)	-0.340 (4.291)	2.155 (5.086)	1.220 (4.017)	1.139 (5.674)	-1.160 (3.789)
Female	4.454 (7.973)	7.964 (8.073)	4.052 (8.165)	2.075 (9.101)	4.734 (9.121)	2.051 (7.212)	6.550 (9.877)	-3.528 (7.021)
Parental Education	0.444** (0.182)	0.431** (0.176)	0.491*** (0.183)	0.414** (0.176)	0.403** (0.204)	0.350** (0.158)	0.389* (0.201)	0.337** (0.143)
Age	-0.445 (0.367)	-0.237 (0.360)	-0.238 (0.370)	-0.553 (0.363)	-0.746* (0.438)	-1.069* (0.563)	-0.509 (0.417)	-0.252 (0.283)
Fail	-2.388 (5.428)	-4.201 (5.408)	-3.024 (5.406)	-2.112 (5.222)	-2.007 (6.160)	0.435 (5.308)	-5.774 (7.704)	-3.091 (4.150)
<i>Variance</i>								
Achievement	0.005*** (0.001)	0.006*** (0.001)	0.005*** (0.001)	0.005*** (0.001)	0.005*** (0.001)	0.005*** (0.001)	0.007*** (0.002)	0.006*** (0.001)
Socioeconomic Index	-0.159 (0.353)	-0.258 (0.360)	-0.124 (0.340)	-0.112 (0.335)	-0.144 (0.382)	-0.052 (0.305)	-0.179 (0.404)	0.139 (0.263)
Age	-0.605 (0.519)	-0.316 (0.507)	-0.379 (0.514)	-0.667 (0.506)	-0.751 (0.578)	-0.837* (0.484)	-0.595 (0.560)	-0.395 (0.372)
Parental Education	0.118 (0.188)	0.137 (0.184)	0.147 (0.185)	0.115 (0.182)	0.059 (0.206)	0.020 (0.170)	-0.040 (0.215)	0.043 (0.143)
N	12,266	11,956	11,507	12,172	11,693	12,135	10,251	8,534

Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . 0: *Baseline*. I-VII: Regressões sem os extremos da distribuição marginal de pré-escola, indicador socioeconômico, branco, dummy de gênero, educação dos pais, idade e reprovação; respectivamente. Todas as especificações incluem os seguintes controles: indicadores de qualidade, de violência, salário do professor, as dummies de urbano, capital, escola privada, estadual, municipal, além das características do professor e diretor.

Tabela 4.22: Análise de Influência - Modelo Estendido -  $\tau = 0.90$ 

$\tau = 0.90$	0	I	II	III	IV	V	VI	VII
$\beta_3$	0.375** (0.164)	0.388** (0.171)	0.435*** (0.140)	0.430*** (0.145)	0.432*** (0.148)	0.357** (0.155)	0.300 (0.208)	0.080 (0.438)
<i>Contextual Peer Effects</i>								
<i>Linear</i>								
Socioeconomic Index	1.346 (0.985)	1.213 (1.043)	1.002 (0.915)	1.021 (0.889)	0.865 (0.880)	1.531 (0.935)	1.880 (1.254)	2.387 (2.194)
White	2.487 (4.472)	2.051 (4.707)	-0.735 (5.388)	2.281 (4.254)	0.772 (4.291)	1.358 (4.130)	3.715 (5.438)	-1.964 (9.154)
Female	15.244* (8.492)	12.293 (8.638)	12.943 (8.428)	12.626 (9.375)	11.770 (8.196)	14.984* (7.800)	13.632 (10.388)	-2.175 (20.537)
Parental Education	-1.925 (2.067)	-2.466 (2.188)	-1.813 (2.079)	-2.055 (1.999)	-2.488 (2.117)	-1.629 (1.954)	1.138 (2.616)	-2.148 (3.936)
Age	9.216 (8.948)	6.282 (9.581)	5.254 (9.307)	12.151 (8.778)	11.223 (9.281)	1.185 (13.010)	11.968 (10.306)	9.016 (16.167)
Fail	6.422 (5.728)	8.694 (5.678)	9.664* (5.220)	8.454 (5.190)	9.387* (5.240)	7.392 (5.344)	2.386 (8.236)	9.081 (10.812)
<i>Squared</i>								
Socioeconomic Index	0.732*** (0.183)	0.796*** (0.217)	0.762*** (0.180)	0.692*** (0.176)	0.541*** (0.178)	0.721*** (0.169)	0.570*** (0.211)	0.511 (0.357)
White	-2.082 (4.528)	-2.385 (4.792)	-0.325 (5.728)	-1.635 (4.304)	0.110 (4.437)	-0.757 (4.239)	-0.794 (5.807)	3.860 (10.520)
Female	-5.986 (8.223)	-3.130 (8.410)	-3.903 (8.163)	-4.455 (9.079)	-2.632 (7.963)	-6.621 (7.582)	-1.809 (9.956)	13.532 (20.701)
Parental Education	0.289 (0.182)	0.313 (0.195)	0.268 (0.182)	0.283 (0.177)	0.328* (0.190)	0.248 (0.171)	0.005 (0.217)	0.341 (0.372)
Age	-0.326 (0.380)	-0.225 (0.406)	-0.166 (0.397)	-0.459 (0.374)	-0.418 (0.396)	0.023 (0.582)	-0.434 (0.437)	-0.350 (0.679)
Fail	-0.109 (5.506)	-3.210 (5.842)	0.257 (5.405)	-1.145 (5.232)	-2.088 (5.463)	-0.475 (5.612)	5.640 (8.109)	-2.909 (11.399)
<i>Variance</i>								
Achievement	0.015*** (0.004)	0.015*** (0.004)	0.014*** (0.004)	0.014*** (0.004)	0.014*** (0.004)	0.016*** (0.004)	0.019*** (0.006)	0.026** (0.012)
Socioeconomic Index	0.217 (0.328)	0.236 (0.344)	0.267 (0.318)	0.224 (0.308)	0.220 (0.317)	0.294 (0.301)	-0.070 (0.400)	0.570 (0.696)
Age	-0.598 (0.532)	-0.473 (0.562)	-0.523 (0.524)	-0.655 (0.512)	-0.663 (0.525)	-1.004* (0.525)	-0.752 (0.598)	-0.874 (1.014)
Parental Education	0.095 (0.199)	0.119 (0.209)	0.201 (0.196)	0.102 (0.191)	0.109 (0.194)	0.040 (0.185)	-0.131 (0.237)	0.198 (0.365)
N	12,266	11,956	11,507	12,172	11,693	12,135	10,251	8,534

Legenda: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . 0: *Baseline*. I-VII: Regressões sem os extremos da distribuição marginal de pré-escola, indicador socioeconômico, branco, dummy de gênero, educação dos pais, idade e reprovação; respectivamente. Todas as especificações incluem os seguintes controles: indicadores de qualidade, de violência, salário do professor, as dummies de urbano, capital, escola privada, estadual, municipal, além das características do professor e diretor.

## Capítulo 5

# Considerações Finais

Neste trabalho investigou-se a natureza e magnitude dos *peer effects* na proficiência dos alunos da quarta-série do ensino fundamental no Brasil. Os resultados apontam que *peer effects* são parte importante dos determinantes do desempenho dos alunos, tanto por meio da composição das características dos pares quanto através do desempenho deles. Dentre as características dos pares, os resultados apontam que o indicador socioeconômico, raça, gênero e idade apresentaram efeitos contextuais significativos na especificação linear do modelo. Nossas estimativas de *peer effects* endógenos ficaram aproximadamente no intervalo [0.25-0.55], evidenciando a presença de um comportamento de conformidade com a norma do grupo nos alunos estudados. Estes resultados implicam num *social multiplier* no intervalo de [1.3-2].

Nós flexibilizamos o modelo *linear-in-means* com o objetivo de verificar as diferenças nas respostas dos alunos conforme o nível das covariadas (efeitos de segunda ordem), além de permitir que estes respondam também a mudanças na variabilidade das características do grupo. Nossos resultados suportam a hipótese de que os alunos em média se saem ligeiramente melhor em turmas mais heterogêneas em termos de habilidade. No entanto, verificamos que este efeito médio camufla uma interessante heterogeneidade, que são justamente os alunos com pior habilidade que saem prejudicados em turmas

heterogêneas, de modo que os benefícios se concentram no centro e no topo desta distribuição. Também verificamos a presença de um efeito médio perverso da variabilidade etária no desempenho dos alunos <sup>1</sup>.

Além disto, verificamos que os *peer effects* endógenos são em geral maiores no topo da distribuição de habilidade, em relação ao centro <sup>2</sup>. Isto significa que os melhores alunos são mais beneficiados por turmas melhores do que trazem benefícios para elas. Esta evidência vai ao encontro da hipótese conhecida como “*the bad apple*”. Esperamos que estes resultados lancem mais luz sobre o debate dos determinantes do desempenho escolar no Brasil, colocando a exploração dos *peer effects* como um mecanismo capaz tanto aumentar o desempenho quanto reduzir a sua desigualdade no país.

---

<sup>1</sup> A dispersão das demais variáveis, como indicador socioeconômico e educação dos pais, não se mostrou relevante na determinação do desempenho dos alunos.

<sup>2</sup> Em relação à base encontramos evidência mista. No modelo *linear-in-means* mais simples verificamos um efeito menor do que no quantil 0.9. Porém, na especificação incluindo termos de segunda ordem e variâncias o efeito encontrado para o quantil 0.10 foi maior do que para o quantil 0.9

# Referências Bibliográficas

- Akerlof, G. A. (1997). “Social distance and social decisions”. *Econometrica*, 65:1005–1027.
- Albernaz, Ângela, F. F. H. G. . F. C. (2002). “Qualidade e equidade no ensino fundamental brasileiro”. *Pesquisa e Planejamento Econômico*, 32:453–476.
- Ammermueller, Andreas; Pischke, J.-S. (2006). “Peer effects in european primary schools: Evidence from PIRLS”. IZA Discussion Paper.
- Angrist, Joshua D.; Graddy, K. I. G. W. (2000). “The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish”. *Review of Economics Studies*, 67 (3):499–527.
- Arnott, Richard; Rowse, J. (1987). “Peer group effects and educational attainment”. *Journal of Public Economics*, 32:287–305.
- Barbosa, Maria Eugênia Ferrão; Fernandes, C. (2001). “A escola brasileira faz diferença? uma investigação dos efeitos da escola na proficiência em matemática dos alunos da 4ª série.”. In *Promoção, ciclos e avaliação educacional*. Org: Franco, Crespo.
- Bishop, John H.; Bishop, M. G. L. G. S. Z. A. (2003). “Nerds and freaks: A theory of student culture and norms”. Brookings Papers on Education Policy.

- Blume, L. E. . D. S. N. (2005). “Identifying social interactions: a review”. Working paper.
- Brock, William A. ; Durlauf, S. N. (2001). “*Interaction-based Models*”, *Hand-book of Econometrics*”, chapter 5, pages 3297–3380. J. Heckman e & E. Leamer.
- Calvo-Armengol, Antoni; Patacchini, E. Z. Y. (2009). “Peer effects and social networks in education”. Mimeo.
- Chatterjee, Samprit; Hadi, A. S. (1986). “Influential observations, high-leverage points and outliers in linear regression”. *Statistical Science*, 1:379–416.
- Coleman, J. S. e. a. (1966). “Equality of educational opportunity.”.
- Cooley, J. (2007). “Desegregation and the achievement gap: Do diverse of peers help?”. Mimeo.
- Durlauf, S. N. (2002). “Groups, social influences and inequality: A memberships theory perspective on poverty traps”. SSRI Working Paper.
- Fletcher, P. (1998). “À procura do ensino eficaz”. Technical report, Ministério da Educação e Cultura - Avaliação da Educação Básica.
- Franco, Creso, O. I. e. a. (2004). “Qualidade e equidade em educação: Re-considerando o significado dos fatores intra-escolares”. Mimeo.
- Graham, B. S. (2005a). “Identifying social interactions through excess variance contrasts”. Working Paper.
- Graham, Bryam S; Imbens, G. W. R. G. (2008). “Measuring the average outcome and inequality effects of segregation in the presence of social spillovers”. Working Paper.
- Graham, Bryan S.; Hahn, J. (2005b). “Identification and estimation of the linear-in-means model of social interactions”. *Economic Letters*, 88:1–6.

- Hanushek, E. A. (1971). “Teacher characteristics and gains in student achievement: estimation using micro data”. *American Economic Review*, 60 (2):280–288.
- Hanushek, Eric A.; Wößmann, L. (2007). “The role of education quality in economic growth”. World Bank Policy Research Working Paper.
- Hausman, J. A.; Taylor, W. E. (1981). “Panel data and unobservable individual effects”. *Econometrica*, 49:1377–1398.
- Hoxby, C. (2000). “Peer effects in the classroom: Learning from gender and race variation”. NBER Working Paper.
- Hoxby, C. (2002). “The power of peers: How does the makeup of a classroom influence achievement?”. *Education Next*, 2:57–63.
- Hoxby, Caroline ; Weingarth, G. (2006). “Taking race out of the equation: School reassignment and the structure of peer effects”. Working Paper.
- Imbens, Guido W.; Newey, W. K. (2003). “Identification and estimation of triangular simultaneous equations models without additivity”. Working Paper.
- Lavy, Victor, D. P. and Schlosser, A. (2007). “Inside the black box of ability peer effects: Evidence from the variation in high and low achievers in the classroom”. Working Paper.
- Ljungqvist, L. (1993). “Economic underdevelopment: The case of missing market for human capital”. *Journal of Development Economics*, 40,:219239.
- Lucas, R. E. (1988). “On the mechanics of economic development”. *Journal of Monetary Economics*, 22:3–42.
- Manski, C. F. (1993). “Identification of endogenous social effects: The reflection problem”. *The Review of Economic Studies*, 60:531–542.

- Manski, C. F. (1995). “Identification problems in the social sciences”. Cambridge, MA: Harvard University Press.
- Marshall, A. (1982). “*Princípios de economia*”. Coleção Os Economistas.
- Mincer, J. (1958). “Investment in human capital and personal income distribution”. *Journal of Political Economy*, 66:281–302.
- Pena, Daniel; Yohai, V. J. (1995). “The detection of influential subsets in linear regression by using an influence matrix”. *Journal of the Royal Statistical Society*, 57:145–156.
- Pinto, C. C. d. X. (2008). “Semiparametric estimation of peer effect in classrooms: Evidence for Brazilian schools in 2003”. Mimeo.
- Sacerdote, B. (2001). “Peer effects with random assignment results for Dartmouth roommates”. *The Quarterly Journal of Economics*, 116:681–704.
- Scheinkman, J. A. (2006). “Social Interactions”. Princeton University and NBER.
- Smith, L. I. (2002). “A tutorial on principal component analysis”. Mimeo.
- UNESCO (2010). “Education for all global monitoring report”. Paris: UNESCO.
- Vygotsky, L. (1978). “Interaction between learning and development”. In *Mind and Society*, chapter 4, pages 79–91. Harvard University Press.
- Wooldridge, J. M. (2002). “*Econometric Analysis of Cross-Section and Panel Data*”. 1st ed. Cambridge, MA.



# Apêndice A

## Primeiro apêndice

### A.1 Construção dos Indicadores

Os indicadores de qualidade e violência na escola, bem como o indicador socioeconômico do aluno foram construídos por meio da análise de componente principal. Não será aqui feita uma revisão desta técnica <sup>1</sup>. Deste modo iremos apenas explicitar quais variáveis foram utilizadas para a construção de cada um dos indicadores.

Para o indicador socioeconômico, utilizamos a informação referente ao consumo de bens duráveis no domicílio do aluno, como o número de carros, banheiros, quartos (por pessoa), TVs, rádios, a presença de aparelho de DVD, geladeira, máquina de lavar, aspirador de pó e, finalmente, a frequência semanal que a família utiliza dos serviços de domésticas.

O indicador de violência na escola utilizou das seguintes informações: presença de tráfico de drogas na escola (agente causador interno), presença de membros da comunidade portando armas (brancas ou de fogo); ações de gangues nas dependências (externas ou internas) da escola e, finalmente, se houve atentado a vida de professores ou funcionários dentro da escola (causado por agente externo).

---

<sup>1</sup>Interessados podem encontrar um excelente guia em Smith (2002)

Por último, para o indicador de qualidade utilizamos da informação sobre o estado de conservação do telhado, paredes, portas, janelas e piso. Além disto, utilizamos também as informações sobre a limpeza da entrada do prédio e das portas <sup>2</sup>.

---

<sup>2</sup> O questionário do SAEB dispõe de uma ampla lista de informações semelhantes possíveis de serem utilizadas para a construção do indicador de qualidade da escola, tais como a limpeza das salas de aula, cozinha, presença de sinais de depredação, etc. Nós utilizamos este conjunto menor para minimizar a quantidade de observações que se perdem devido a *missing data*. O uso de um vetor menor não parece acarretar problemas tendo em vista que a correlação entre o indicador utilizado e o composto por mais de 30 itens do questionário foi superior a 0.8. Esta nota também se aplica ao indicador de violência.