



Universidade de Brasília

Departamento de Economia

Dissertação de Mestrado

O Impacto da Infraestrutura Escolar no Rendimento dos Alunos

Brasília/DF

2014

O Impacto da Infraestrutura Escolar no Rendimento dos Alunos

Daniel Góes Cavalcante

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciências Econômicas pela Universidade de Brasília.

Aprovada pela Banca Examinadora:

Prof. Moisés de Andrade Resende Filho (Orientador)

Prof. Roberto Góes Ellery Júnior (Membro Interno)

Dr. Pedro Erik Arruda Carneiro (Membro Externo)

Brasília

Agosto de 2014

RESUMO

A relação entre os recursos escolares e o rendimento escolar dos alunos é bastante controversa na literatura, em grande parte porque envolve políticas públicas tradicionais voltadas para a melhora do capital humano social. No que se restringem ao impacto das infraestruturas física e organizacional das escolas as evidências também indicam pouca consistência na sua importância no desempenho dos alunos. Portanto, o objetivo do trabalho presente é estimar o impacto da infraestrutura na proficiência em matemática dos alunos de escolas públicas do Brasil fornecidos pelo Sistema de Avaliação da Educação Básica (Saeb) do ano de 2011. Para tanto, em vez utilizar um insumo de infraestrutura em específico como variável de interesse foi preferido utilizar um indicador de infraestrutura calculado para este trabalho envolvendo diversas variáveis ligadas a infraestrutura. Além disso, deseja-se entender como a infraestrutura afeta os estudantes em seus diferentes percentis de distribuição na proficiência. Com esses objetivos, os modelos empregados são o de Mínimos Quadrados Ordinários e o de Regressão Quantílica. Os resultados indicam que o impacto da infraestrutura é positivo e significativo. Além disso, o efeito não é homogêneo ao longo da distribuição da nota de proficiência.

Palavras-Chave: Resultados Educacionais, Infraestrutura Escolar, Regressão Quantílica, Função de Produção Educacional.

ABSTRACT

The relationship between school resources and the students achievement has been controversial on literature, in large because it involves questions about a variety of traditional policies in human capital approach. When the focus is restricted to the impact of the physical and organizational infrastructure on student's achievement, evidences also shows little support for them. Therefore, instead of using a specific infrastructure input was preferably using an indicator of infrastructure calculated for this work involving a variety of others variables of infrastructure. Furthermore, it wanted to understand how infrastructure affects the student performance in the different degrees of the proficiency distribution. With these objectives, the models employed here were the Ordinary Less Squares and Quantiles Regressions. The results indicate that the impact of educational infrastructure is positive and significant. Moreover, the effect is not the same for all the distributions in proficiency.

Key-Words: Educational Results, School Infrastructure, Quantile Regression, Educational Production Function.

Sumário Executivo

INTRODUÇÃO	1
FUNÇÃO DE PRODUÇÃO EDUCACIONAL.....	3
FORMA FUNCIONAL DA FUNÇÃO DE PRODUÇÃO EDUCACIONAL.....	4
PROBLEMAS COM RELAÇÃO À FUNÇÃO DE PRODUÇÃO EDUCACIONAL .	8
MEDINDO O PRODUTO	9
A IMPORTÂNCIA DOS INSUMOS E EM ESPECIAL DA INFRAESTRUTURA	10
METODOLOGIA E BASE DE DADOS.....	15
BASE DE DADOS.....	15
METODOLOGIA	24
A ABORDAGEM DA FUNÇÃO DE PRODUÇÃO EDUCACIONAL NO MODELO DE MÍNIMOS QUADRADOS ORDINÁRIOS (MQO)	25
A ABORDAGEM DA FUNÇÃO DE PRODUÇÃO EDUCACIONAL NO MODELO DE REGRESSÃO QUANTÍLICA	28
ANÁLISE DOS RESULTADOS	31
CONCLUSÃO	41
REFERÊNCIA BIBLIOGRÁFICA.....	43

Lista de Tabelas

IMPACTO DA INFRAESTRUTURA FÍSICA DA ESCOLA NA NOTA EM TESTES.....	13
IMPACTO DA INFRAESTRUTURA ORGANIZACIONAL NA NOTA EM TESTES.....	14
DISTRIBUIÇÃO DE PROFICIÊNCIA EM MATEMÁTICA.....	18
CARACTERÍSTICAS DOS ALUNOS SEGUNDO A DISTRIBUIÇÃO DAS PROFICIÊNCIAS DE MATEMÁTICA-QUINTA SÉRIE	18
CARACTERÍSTICAS DOS ALUNOS SEGUNDO A DISTRIBUIÇÃO DAS PROFICIÊNCIAS DE MATEMÁTICA-NONA SÉRIE.....	18
CARACTERÍSTICAS DOS ALUNOS SEGUNDO A DISTRIBUIÇÃO DAS PROFICIÊNCIAS DE MATEMÁTICA-TERCEIRO ANO.....	19
IMPACTO MÉDIO DA PROFICIÊNCIA EM MATEMÁTICA-5ª SÉRIE	33
IMPACTO MÉDIO DA PROFICIÊNCIA EM MATEMÁTICA-9ª SÉRIE	34
IMPACTO MÉDIO DA PROFICIÊNCIA EM MATEMÁTICA-3º SÉRIE	35
TESTE BREUSCH-PAGAN.....	36
IMPACTO DA INFRAESTRUTURA DA ESCOLA NAS NOTAS DE ACORDO COM OS QUANTIS DE DISTRIBUIÇÃO-5ª SÉRIE.....	39
IMPACTO DA INFRAESTRUTURA DA ESCOLA NAS NOTAS DE ACORDO COM OS QUANTIS DE DISTRIBUIÇÃO-9ª SÉRIE.....	39
IMPACTO DA INFRAESTRUTURA DA ESCOLA NAS NOTAS DE ACORDO COM OS QUANTIS DE DISTRIBUIÇÃO-3º ANO.....	39

Lista de Quadros

**DESCRIÇÃO DAS VARIÁVEIS UTILIZADAS PARA A MONTAGEM DO
INDICADOR DE INFRAESTRUTURA..... 22**

DESCRIÇÃO DAS VARIÁVEIS UTILIZADAS NO MODELO 24

Lista de Figuras

IMPACTO DA INFRAESTRUTURA FÍSICA DA ESCOLA NA NOTA EM TESTES.....	40
IMPACTO DA INFRAESTRUTURA ORGANIZACIONAL NA NOTA EM TESTES.....	40
DISTRIBUIÇÃO DE PROFICIÊNCIA EM MATEMÁTICA.....	41

1. Introdução

Economistas têm dado muita atenção em entender como o capital humano afeta desde indicadores sociais e urbanos até o crescimento do produto. A primeira noção de capital humano presume que os indivíduos tomam decisões de investir neles mesmos com educação e, a partir disso, as habilidades acumuladas ao longo do tempo e que são importantes para o mercado passam a fazer parte do capital humano individual. Os investimentos feitos na melhoria de suas capacidades retornam em benefícios assim como uma empresa investe em máquinas para aumentar a produção.

No que se refere ao capital humano e no impacto desse componente no crescimento econômico, economistas são céticos sobre qualquer afirmação exata de onde e como as habilidades são adquiridas. Entretanto, é algo comum presumir que a educação formal é um dos mais importantes contribuintes para a melhoria das habilidades pessoais e o aumento do capital humano. Apesar de outros fatores serem tão ou mais importante do que a educação formal, como o envolvimento dos pais, as habilidade próprias e a influência de colegas, as escolas são especiais porque são diretamente afetadas por políticas públicas.

A maior parte dos trabalhos empíricos e das políticas públicas voltadas à educação já realizado se preocupou principalmente na quantidade da educação fornecida. Todavia, com o aumento por todos os continentes da maior participação das crianças e adolescentes nas escolas o foco mudou para o entendimento das atribuições das escolas e de como a qualidade do repasse da educação das crianças refletem nas habilidades dos estudantes e no aumento do capital humano¹.

Muitos economistas tendem a enfatizar os resultados no mercado de trabalho quando se pensa em diferenças nas habilidades individuais. Entretanto, no que concerne a discussões com relação à educação, estudos voltados diretamente aos efeitos das habilidades individuais no mercado de trabalho não possuem sentido prático, porque esses efeitos serão sentidos muitos anos depois.

A alternativa largamente utilizada considera, ao menos conceitualmente, um procedimento de duas etapas. O primeiro passo envolve a consideração de como escolas e outras influências em estudantes se relacionam com uma *proxy* das habilidades individuais, *proxy* esta que geralmente se relaciona com habilidades cognitivas retiradas em testes de proficiência. O segundo passo considera como essa *proxy* se relaciona com os resultados no mercado de trabalho (Hanushek, 2006). O segundo passo é menos frequentemente estudado na literatura e de certo modo é apenas assumido como verdadeiro.

¹ Segundo dados do Banco Mundial, as matrículas do ensino secundário aumentaram consideravelmente de 1990 até 2011, principalmente na América Latina, que viu sua taxa líquida de matrícula (aquela que se refere aos alunos com idade para estar na série designada) subir de 49,35% para 76,14%.

Obviamente, habilidades cognitivas não são as únicas variáveis importantes para os resultados econômicos de interesse, apesar de ser bastante consensual de que elas afetam positivamente os resultados no mercado de trabalho. Outras variáveis são tão importantes quanto, como habilidade em trabalho em equipe, conhecimento político e criatividade. Além disso, variações do grau de habilidade cognitiva capturadas pelos testes de proficiência são também influenciadas por fatores associados às características da família, dos colegas e da vizinhança. O fato é que os estudantes e os pais se preocupam especialmente com a qualidade da educação oferecida pelas escolas e levam isso em consideração ao tomar diversas outras decisões.

Mensurar a qualidade das escolas e mostrar a relação entre ela e o desenvolvimento cognitivo dos alunos não é algo simples e para isso são utilizadas medidas que se supõe ter algum grau de relação com a qualidade das escolas, apesar de que as variáveis comumente utilizadas na literatura, como gasto por aluno e proporção aluno-professor, serem conhecidas por *proxies* pobres da qualidade da escola. (Hanushek, 1992).

Variáveis de infraestrutura também são largamente utilizadas na literatura com a finalidade de mostrar a importância, ou não, da qualidade das escolas e os resultados em geral são ambíguos. Leigh L. Linden (2008) estuda o efeito de um programa computacional que ajuda no aprendizado de assistência em matemática e observa que os alunos que tiveram acesso ao programa apresentaram resultados piores em 0.57 desvios padrões do que um grupo de controle que não teve acesso ao programa. Por outro lado, Gleew et.all (2004) avalia o impacto do uso de apresentações de slides nas notas de alunos em escolas no Quênia e o efeito é positivo, apesar de ser pequeno.

Os recursos educacionais disponíveis também são importantes e são mais ligados a infraestrutura organizacional do sistema e também da escola. Após controlar o PIB per capita, 33% da variação nas notas em matemática entre os países da OCDE, podem ser explicadas por diferenças nas respostas de diretores sobre a adequação dos equipamentos de laboratório, materiais instrucionais (livros textos, entre outros), uso de computadores, conexão a internet, programas computacionais como acessório, e biblioteca (PISA Volume IV, 2012).

Esse trabalho visa mostrar como a infraestrutura impacta o rendimento dos alunos em exames de proficiência em matemática. Para isso, ao invés de utilizar uma variável única como mensuração da infraestrutura disponível nas escolas, foi usado um indicador de infraestrutura montado a partir de diversas variáveis ligadas a infraestrutura física e organizacional das escolas, dentre elas a disponibilidade de computadores e internet nas escolas, o estado das salas de aula e o acesso a bibliotecas e laboratórios. A vantagem de utilizar um indicador de infraestrutura é que ele fornece a possibilidade de entender como os alunos reagem ao aumento da infraestrutura como um todo, já que um componente de infraestrutura específico podem impactar de forma diferente os estudantes de acordo com a associação dele com outros componentes de infraestrutura.

Os resultados são apresentados em duas etapas. A primeira analisa o efeito do indicador em dados *cross-sectional* no modelo MQO padrão robusto, controlando também para variáveis ligadas ao *background* familiar, características dos alunos e regiões geográficas. A segunda etapa decorre da busca por entender como os estudantes reagem ao indicador de

infraestrutura de acordo com as posições deles ao longo da distribuição de notas dos alunos. Para isso o modelo utilizado será o de regressão quantílica. Espera-se que os resultados mostrem uma relação positiva entre o indicador e a nota de proficiência dos alunos e que o poder de influência dele seja maior quanto menor for a nota do aluno. A base de dados utilizada para as estimações é o Sistema de Avaliação da Educação Básica (Saeb) de 2011.

O capítulo 2 será uma resenha bibliográfica focada principalmente na problemática da função de produção. No capítulo 3 será explicitado como é construída a base de dados do Saeb, as variáveis utilizadas nas regressões do trabalho, assim como a montagem do indicador de infraestrutura e os modelos empíricos. No capítulo 4 serão mostrados os resultados dos modelos em MQO e os modelos de regressão quantílica. Por fim, no capítulo 5 a conclusão do trabalho.

2. Função de Produção Educacional

No curso de microeconomia o estudante de economia se depara com o conhecimento da função de produção, usada principalmente para analisar a decisão da firma. Em sua representação abstrata, as possibilidades de produção da firma são obtidas através de uma variedade de relações tecnológicas. A função de produção descreve o máximo de produção possível dada a quantidade de insumos utilizados (Monk, 1989). A firma tem como sua função “natural” utilizar os recursos disponíveis a fim de maximizar os resultados. Basicamente é esta a motivação do estudo da função de produção².

Na literatura da função de produção educacional, pesquisadores tentam desenhar uma analogia entre o processo de aquisição e acúmulo de conhecimento e o processo de produção de uma firma. O primeiro objetivo do pesquisador é entender a tecnologia que combina os insumos das escolas e que cria resultados cognitivos dos alunos (Wolpin, 2003). A função de produção análoga à que se obtém tradicionalmente na indústria disponibiliza um quadro que guia a escolha das variáveis e permite a interpretação de seus efeitos.

No estudo sobre as causas do desempenho educacional, a relação de insumo-produto ganhou destaque na literatura econômica nas últimas décadas. A atenção dada à relação de insumos e produto na educação vislumbra medidas para melhorar a gestão das escolas, obter informações sobre políticas de acesso a educação, políticas de integração escolar e financiamento escolar.

² ¹ Adicionalmente, poderíamos tratar aqui de algumas considerações da função de produção e da relação de produção. Primeiro, poderíamos estendê-la para análise de múltiplas variáveis no produto. Segundo, no curso de microeconomia, o produto marginal dos insumos é sempre decrescente, o que significa que o uso maior de um único insumo trás retornos cada vez menores à produção. Terceiro, há insumos que são complementares a outros insumos, ou mesmo substitutos.

O primeiro e um dos mais importantes trabalhos na literatura educacional utilizando função de produção foi o *Equality of educational opportunity* ou *Coleman Report*. Este foi um estudo encomendado pelo governo americano (seção 402 do *Civil Right Act* de 1964) com o objetivo primordial de entender a desigualdade de notas e avaliações escolares entre grupos de cores e grupos étnicos no ensino primário e secundário nos Estados Unidos, sendo importante por diversos motivos. Primeiro, a base de dados coletada obtinha informações de mais de 600 mil alunos de mais de 3 mil escolas de todo os Estados Unidos. Informações das escolas também foram observadas na análise. Segundo, ele direcionou a atenção a estudos sobre a relação entre insumos escolares e desempenho escolar dos alunos.

O relatório indicou que o desempenho dos alunos estava basicamente ligado ao perfil da família e de sua vizinhança. Segundo ele, famílias mais ricas, de maior escolaridade e que participavam mais da vida estudantil da criança geravam quase que todas as condições para o bom rendimento escolar dos alunos.

O conceito de função de produção educacional não é totalmente similar à função de produção dada nos livros textos de microeconomia, que consideraria escolas como “fábricas” que produziriam “conhecimento” usando várias características das escolas e dos professores como insumos. A razão para isso é que o processo pelo qual as habilidades cognitivas são adquiridas envolve uma imensidão de fatores, sendo funções de produção apenas expressões desse processo de “aprendizado”. Dessa forma, apesar de ser factível encontrar uma função de produção educacional que descreva perfeitamente a relação entre os insumos escolares e o resultado dos alunos, essa relação é quase impossível de ser achada. Se isso acontecesse, então seria possível encontrar as combinações das características das escolas e das características dos professores que, somando-se informações sobre os seus custos, levem ao resultado desejado da maneira mais efetiva possível, dada a restrição orçamentária.

2.1. Forma Funcional da Função de Produção Educacional

Para entender os resultados dos estudos que serão mais a frente apresentados e a forma funcional da função de produção é preciso dar um passo atrás e entender quais as relações de interesse e como essas relações são atreladas com o comportamento familiar.

Diferentemente da teoria da firma, em que o processo de decisão se baseia na escolha da combinação de insumos que darão a resposta ótima, a perspectiva educacional acarreta o entrelaçamento entre as decisões dos estudantes e dos pais com a decisão das escolas.

Para entender melhor essa relação, assumamos que os pais maximizam uma função de utilidade da família temporal sujeito às restrições impostas. A função de utilidade familiar seria função do consumo de bens e serviços, incluindo lazer, e da quantidade de educação dos filhos. A restrição encarada pelos pais envolve o orçamento familiar no tempo, o impacto dos anos de estudo na expectativa de renda futura dos filhos e uma função de produção de aprendizado dos pais.

Dessa forma, seguindo o raciocínio de Glewwe e Kremer (2006), a função de produção de aprendizado dos filhos é:

$$A = a(S, Q, C, H, I) \quad (2.1)$$

Em que A são as habilidades aprendidas, S é anos de estudo, Q é um vetor de características da escola e dos professores, C é um vetor de características dos alunos (habilidades inatas), H é um vetor das características domésticas. I é um vetor de insumos escolares no controle dos pais, tais como livros didáticos e materiais de apoio.

Suponha que os vetores C e H são exógenos, isto é, não podem ser mudados pelos pais e nem pelos estudantes. Outras características das crianças que podem afetar no seu rendimento, como o estado de saúde, são endógenos (controlados pelos pais) e podem fazer parte do vetor I .

Em um cenário simples, os pais tem acesso apenas a uma única escola e as características dela não são influenciadas pelos pais. Portanto, todas as variáveis em Q são exógenas a H . Os pais podem escolher S e I (sujeito às restrições mencionadas) para maximizar a função de utilidade da família. Isso implica que anos de estudo S e os insumos da escola podem ser representados por:

$$S = f(Q, H, I, P) \quad (2.2)$$

$$I = g(Q, C, H, P) \quad (2.3)$$

Em que P denota o vetor de preços em relação à escola (mensalidade, livros, outras taxas, uniformes), que também é exógeno. Inserindo as equações (2.2) e (2.3) em (2.1), obtemos:

$$A = h(Q, C, H, P) \quad (2.4)$$

Essa forma reduzida não estabelece uma relação causal e não pode ser relacionada às funções de produção usuais porque reflete preferências familiares e inclui preços no seu argumento.

Um cenário mais realista seria a de que os pais teriam mais do que uma única opção de escola para colocar os filhos, e por isso Q e P seriam endógenos. Dessa forma, na maximização da utilidade da família, os pais escolhem primeiro as escolas. Daí, condicional à escolha da escola, os pais então escolhem S e I .

A política educacional em geral tem interesse no impacto da característica das escolas e dos professores (Q), nos preços relacionados à educação na escola (P), nos anos de estudo de sua população (S) e no rendimento acadêmico dos estudantes (A). Como exemplo, aumentar a qualificação dos professores alteraria Q e subsidiar o preço da mensalidade das escolas pode ser visto como uma alteração em P . As equações (2.2) e (4) mostram como uma alteração em P afeta S e A . A equação (2.2) também mostra como uma alteração em Q afeta a quantidade de anos de estudo, S .

Voltando atenção para o impacto da infraestrutura escolar no rendimento dos estudantes, suponha que se escolha uma variável Q relacionada à qualidade da escola (poderia ser o estado de conservação das salas de aula). A equação (2.1) mostra como o impacto de Q afeta A parcialmente, quando todas as outras variáveis permanecem constantes. Por outro lado, a equação (2.4) provê o impacto total em A com respeito à Q , porque permite que sejam feitas alterações em I e S com respeito à Q . Os pais podem ser afetados por aumentos na provisão em infraestrutura positivamente com relação à sua provisão em insumos educacionais, como livros. Se os pais acharem que os insumos escolares agem como substitutos aos insumos fornecidos por eles, o aumento da provisão em infraestrutura escolar afeta negativamente a atenção dos pais.

O fato das ações dos pais poderem reforçar ou mitigar a decisão das escolas ajudam a explicar o porquê das estimativas do impacto dos insumos escolares (como veremos adiante na próxima seção) trazerem muitas vezes resultados inconsistentes. Além disso, diferentes estudos trazem resultados diferentes do impacto dos insumos escolares no rendimento dos alunos porque muitos deles fazem a estimativa da função de produção (equação 2.1), enquanto outras fazem a estimação da forma funcional reduzida (equação 2.4).

Apesar de a forma funcional reduzida parecer ser a mais indicada para proposições políticas, a resposta não é tão clara quanto parece. A equação (2.4) é importante porque ela mostrará o que de fato irá ocorrer no rendimento dos alunos depois de uma mudança na oferta dos insumos escolares. Por outro lado, a equação (2.1) não vai mostrar isso porque ela não captura o efeito da qualidade da escola (Q) ou dos preços (P) na quantidade de anos de estudo (S) ou na oferta de insumos escolares pelos pais (I).

Por sua vez, a equação (2.1) ainda pode ser de interesse, pois ela captura o impacto no bem estar total da família. Se os pais, por exemplo, respondem ao incremento na qualidade da educação fornecido pela escola reduzindo seus gastos com insumos educacionais para o filho, então ele terá mais dinheiro para comprar outros bens para a família, o que gera maior impacto no bem estar familiar. Na equação (2.4), o impacto de Q em A reflete o efeito negativo da diminuição de I em A , mas não leva em consideração um eventual aumento de A por influência positiva do aumento do

consumo familiar. O impacto estrutural medido pela equação (2.1) não considera nenhum efeito. Assim, a equação (2.1) tende a cancelar esses efeitos opostos e fornecer a mensuração mais aproximada da mudança no bem estar familiar com respeito a uma mudança em A .

Voltando atenção para a função de produção, em sua forma mais geral, ela é estabelecida como:

$$A = f(Q, H, C, D) \quad (2.5)$$

Em que A é a medida de desempenho dos alunos, Q envolve características ligadas à qualidade da escola, como infraestrutura e qualificação dos professores, H é ligada ao *background* familiar, C é uma medida que tenta capturar o esforço dos alunos e D é o *peer effect*, ou a característica dos colegas.

Geralmente, os insumos escolares utilizados em análises empíricas incluem medidas objetivas sobre características das escolas e dos professores, como: nível de escolaridade do professor, experiência, condições das salas de aula, suporte de apoio aos alunos (cadernos de apoio e livros didáticos) e tamanho da turma.

Do ponto de vista organizacional da escola, há dois tipos de variáveis com respeito ao processo e a organização. Primeiro, são variáveis observáveis ligadas às características organizacionais e processuais da escola (ou macro variáveis), que representam claramente a reprodução de práticas educacionais, como organização das aulas, currículo e número de períodos disponíveis para os alunos. O segundo tipo são as variáveis que representam as características processuais ligadas às características dos professores (ou micro variáveis), como gerenciamento da classe, métodos de ensinamento e habilidades de comunicação.

Embora haja problemas com as macro variáveis ao captar o real impacto desses macros processos, dado que há uma grande diversidade quanto a forma com que são implementados, ainda há como fazer estimações em funções de produção utilizando essas macro variáveis (uma das mais utilizadas nesse contexto é o tamanho da turma). O problema maior ocorre com as micro variáveis. A maior parte das decisões na questão educacional se dá de modo micro, tomadas por agentes, em sua maioria professores. Essas são dificilmente observáveis, mensuráveis e reproduzíveis. Essas questões micro processuais podem ser interpretadas como diferenças nas habilidades de quem faz parte da estrutura educacional.

O reconhecimento desses dois diferentes níveis de tomada de decisão gera implicações na eficiência de produção³ e altera a interpretação que há com relação aos

³ Uma importante questão é se as escolas são eficientes na função de produção. Isso tem implicações importantes na política escolar, já que ineficiência indica que melhores resultados podem ser alcançados utilizando os mesmos tipos de insumos. Desde que as estimações são baseadas em variáveis

professores e aos insumos escolares. É perfeitamente clara a importância dos atributos mensuráveis dos professores (diploma, nível de qualificação, experiência), já que a maioria das escolas utiliza como referência na hora de contratar ou de estabelecer os salários, esses atributos. No entanto, o impacto estimado dessas variáveis observáveis deve ser utilizado para entender o que está por trás dessas estimações. Por exemplo, como interpretar um efeito negativo do fato dos professores serem pós-graduados sobre os resultados dos alunos? Será mesmo que os professores pós-graduados geram efeito negativo nos alunos, ou na hora de contratar um professor o avaliador leva em conta como aspecto mais relevante o currículo de formação do candidato, sendo que dessa forma há uma ingerência macro processual não observada?

Uma das implicações dessa discussão é que mais esforço deve ser dado no estudo dessas relações macro e micro processual e organizacional das escolas. Até porque, não há nenhuma indicação de que as escolas fazem a escolha pelo melhor processo e organização com respeito aos insumos disponíveis⁴. O ideal na análise do processo e da organização escolar seria a obtenção do máximo de informação útil para que os métodos mais eficientes possam ser replicados em situações diferentes.

2.2. Problemas com Relação à Função de Produção Educacional

Nesta seção será feito breve um relato dos potenciais problemas encontrados nas estimações de funções de produção educacional.

Considere a estimação de uma simples função linear de produção educacional, especificada como:

$$A = \beta_0 + \beta_1 Q + \beta_2 H + \beta_3 C + \beta_4 D + u_a \quad (8)$$

Em que cada vetor dessa equação é representado da mesma forma que a equação (1). O termo erro faz parte da equação por três motivos basicamente. Primeiro, não é possível encaixar no modelo todas as variáveis explicativas ligadas aos vetores Q , H , C e D por falta de dados. Segundo, o termo erro indica que é apenas uma aproximação linear do modelo verdadeiro. Terceiro, a variável explicada, A , pode mensurar as habilidades naquele momento de maneira inexata.

observáveis do comportamento da escola, as relações estimadas talvez não construa a linha da fronteira da produção se as escolas não tiverem produzindo o máximo dado os insumos disponíveis. Nesses casos, as relações descreverão o comportamento médio que não é a maneira apropriada em ver como a adição de insumos afeta no produto.

⁴ Dois argumentos podem ser dados para sustentar a hipótese de que as escolas não são eficientes tecnicamente. O primeiro é que quem toma decisão na educação não é guiado por fundamentos que maximizem os lucros, principalmente se a escola for pública. O outro argumento é de que possivelmente o tomador de decisão não conhece a estrutura do processo educacional, dessa forma não há como esperar que a escola esteja na fronteira da produção. (Hanushek, 1979)

Com relação às variáveis explicativas, podemos listar alguns problemas encontrados. O primeiro problema é relacionado à omissão de variáveis. O impacto da variável observada na aprendizagem, ou o coeficiente β , pode ser consistentemente estimado por Mínimos Quadrados Ordinários se as variáveis que entram no modelo forem não correlacionadas com o erro u_a . Infelizmente, o usual é que haja algum tipo de correlação. Qualquer correlação entre essas variáveis de dentro e de fora do modelo causa viés, sendo esse viés proporcional ao nível de correlação entre essas variáveis.

A segunda preocupação é com a simultaneidade. Esse problema acontece quando o processo de decisão não é um fator isolado. As variáveis dentro do modelo variam em decorrência de outras. Por exemplo, a escolha da escola sobre a alocação dos recursos é influenciada pela preferência dos pais, assim como a preferência dos pais é influenciada pela escolha na alocação dos recursos da escola. Outro exemplo: a didática do professor pode ser dependente também da atenção e da qualidade dos alunos, assim como o oposto é verdadeiro também. Esse problema pode causar má especificação no modelo e causar viés.

O terceiro problema é relacionado à seleção de amostra. Viés de seleção amostral ocorre quando a amostra não é aleatória e um ou mais subgrupos são menos prováveis de serem incluídos na amostra do que outros. Para esclarecer, suponha que se queira estudar a relação entre o desempenho dos alunos na escola e a frequência escolar dos mesmos. Aqui, a variável explicativa é a frequência escolar, enquanto a variável explicada é o desempenho dos alunos. O problema é que pode ser que quem não vai à escola não vai porque a escola não tem condições de fornecer merenda ou de fornecer transporte de casa até a escola. Portanto, os resultados seriam tendenciosos por pegar um grupo que tem condições de não depender da escola para se alimentar ou para se transportar.

O quarto problema é causado por erro de medida nas variáveis. Esse problema tende a gerar um coeficiente se aproximando de zero (viés de atenuação) para as variáveis com erro de medida, tornando o efeito insignificante.

2.3. Medindo o Produto

A maioria dos estudos de função de produção na educação considera como medida de produto os resultados dos alunos em testes padronizados. Outros estudos consideram também atitude dos estudantes (Levin, 1971), taxa de adimplência na escola (Katzman, 1970) e continuação na escola ou taxa de abandono (Katzman, 1970).

Enquanto a teoria da produção tradicional concentra sua análise na variação de insumos em um produto final único ou múltiplo, mas homogêneo, o estudo da função de produção na educação não funciona desse jeito. A educação transforma insumos

muitas vezes dificilmente mensuráveis (a qualidade do professor ou a atenção dos pais) em “estudantes” com qualidades diferentes, por exemplo, com distintas capacidades de absorção de conhecimento. Dessa forma, a imposição de testes de habilidade cognitiva como padrão para a obtenção de resultados com relação aos níveis de aprendizado dos alunos pode não ser a melhor forma de identificar a resposta dos alunos aos insumos.⁵

Card e Krueger (1992a) fazem uma crítica ao modelo de função de produção que utiliza como produto a nota padronizada dos alunos em exames de proficiência argumentando que estes testes muitas vezes avaliam características dos alunos sem valor econômico. Segundo eles, o produto deveria ter como base o sucesso que os estudantes adquirem posteriormente no mercado de trabalho (salário, renda).

No entanto, há uma dificuldade enorme de se inserir informações sobre o perfil acadêmico e a qualidade do ensino do trabalhador em estudos sobre o mercado de trabalho. Um dos trabalhos mais proeminentes e de grande contribuição para a literatura é o trabalho feito por Card e Krueger (1992a). Esse estudo analisa a relação entre a qualidade dos insumos educacionais e os ganhos no mercado de trabalho anos depois. O resultado é uma relação positiva entre a qualidade dos insumos e os ganhos dos estudantes no mercado de trabalho.

Os testes padronizados de habilidades cognitivas possuem a vantagem de serem facilmente aplicáveis e de serem comparáveis numa região, ou mesmo pelo mundo. Ademais, existe uma tendência de que os testes sejam cada vez mais mercado-orientados.

Claramente, o interesse no desempenho do sistema de educação está no interesse em saber como a escola pode influenciar o desenvolvimento de habilidades nos alunos que possam ser aproveitados por ele e pela sociedade após a saída dele da escola. Considera-se aqui o efeito da escola para o aluno no mercado de trabalho e a influência do ensino na socialização do indivíduo.

2.4. A Importância dos Insumos e em Especial da Infraestrutura

⁵ Apesar de haver um aperfeiçoamento dos testes padronizados de modo a identificar as habilidades dos alunos e como eles reagem a determinados insumos disponibilizados, ainda não há uma clara orientação no sentido de saber como e quando as habilidades são desenvolvidas. Por exemplo: pode ser que o maior rendimento do aluno fruto da atenção que este dá ao professor não seja um mérito do professor em chamar esta atenção, mas uma habilidade do estudante, adquirida nos primeiros anos de vida. Sendo que esta habilidade pode ter sido gerada por qualquer motivo, por exemplo, por um motivo completamente diferente daquele que levou um outro estudante a desenvolver esta variável.

Apesar de uma adequada medida do produto para a função de produção ser um tema relevante, o foco dos estudos dessa área foi transferido para os insumos escolares. Isso ocorre principalmente para aqueles insumos relevantes para as políticas públicas, tais como a infraestrutura, qualificação e remuneração dos professores.

Os modelos padrões de função de produção são basicamente resumidos da seguinte forma: o produto do processo educacional (o sucesso do estudante em termos educacionais) é diretamente relacionado aos insumos, que podem ser controlados por políticas públicas (característica das escolas, professores, currículo, gastos por aluno), ou que podem não ser controlados, como é o caso da família, amigos, capacidade de aprendizado do estudante e riqueza.

O “background” familiar geralmente é caracterizado por algumas características sociais e demográficas, tais como educação dos pais, renda e tamanho da família. A influência dos amigos e colegas de classe são geralmente medidas pelas características sociais e demográficas ou pelo sucesso obtido na escola. A característica das escolas é comumente avaliada pela sua infraestrutura, organização e recursos financeiros.

Ao longo dos anos diversos trabalhos foram feitos no sentido de investigar a relação entre os insumos educacionais e o sucesso dos alunos tanto em testes padronizados quanto em outras medidas de alcance. A maioria deles focou o estudo em algumas variáveis, o que de fato é um mérito no sentido de padronizar os estudos e permitir que eles sejam comparados e aperfeiçoados.

A partir dos inúmeros estudos foi possível partir para uma categoria que é comumente chamada de “meta-análise”, que consiste na junção desses trabalhos que tem em comum o mesmo tema de interesse (por exemplo: a importância da razão professor/aluno para o rendimento dos alunos) para então consolidar um resultado padrão obtido. Seu papel principal seria generalizar para uma população mais abrangente os resultados encontrados, mesmo entendendo a heterogeneidade dos efeitos e as diferenças estatísticas metodológicas.

A literatura da meta-análise, porém, carrega um possível problema, que é o viés de publicação. Em particular, muitos autores podem publicar artigos apenas se os resultados encontrados vierem de acordo com o que eles gostariam. Ou então, as editoras poderiam tender a publicar apenas trabalhos com resultados significantes.

Hanushek (2003) em seu artigo “Failure of Input-Based Schooling Policies”, faz um estudo sobre o fracasso das pesquisas empíricas em evidenciar uma relação significativa entre diversos insumos, principalmente aqueles mais utilizados em políticas públicas, e a avaliação do rendimento escolar. O trabalho inclui todas as publicações feitas até 1995 utilizando dados de escolas dos Estados Unidos, que provém medidas sobre características da família e das escolas com relação ao alcance

dos estudantes, e que dispõe dos sinais e das significâncias estatísticas. Foram 89 publicações que disponibilizaram 376 funções de produção estimadas⁶.

Três-quartos das estimações consideram a variável dependente como o resultado dos alunos nos testes de proficiência, enquanto os demais utilizam uma variedade de outras medidas, como: continuação na escola, evasão escolar e ganhos no mercado de trabalho subsequente.⁷ Grande parte dos trabalhos que utilizam outras formas de mensuração do sucesso estudantil diferentes do resultado de testes padronizados encontra-se em estimações feitas para o ensino secundário.

Em termos de recursos escolares, só 9% das estimações consideram o grau de educação dos professores e 14% das estimações investigando a relação professor-aluno encontram efeito positivo e estatisticamente significativo na despesa dos alunos. Respectivamente, 72% e 86% das estimações consideram os sinais encontrados estatisticamente insignificantes. A experiência do professor possui uma relação positiva e estatisticamente significativa em 29% das estimações. Mesmo assim, 68% das estimações são estatisticamente insignificantes. No entanto, há relativamente uma elevada taxa da relação positiva com coeficientes estatisticamente significantes entre resultados nos testes para professores e sucesso dos alunos, 37%.

Com relação a variáveis de custo financeiro, o salário dos professores afeta positivamente com coeficiente estatisticamente significativo os resultados dos alunos em 20% das estimações, enquanto essa relação para ganhos por aluno é de 27%. Sendo que a maioria deles expõe elevados percentuais de dados não significantes.

Entende-se como infraestrutura adequada na escola a disponibilidade de salas de aula com espaço e luminosidade suficientes, devidamente arejadas, instalação hidráulica e elétrica em bom estado, disponibilidade de carteiras para todos os alunos, acesso a livros didáticos entre outras características que deem suporte para que o aluno se desenvolva cognitivamente.

Não há na literatura um consenso sobre a direção do impacto de cada insumo no aprendizado dos alunos, nem da importância relativa de cada um deles. O que se tem é apenas uma noção, com base em evidências empíricas, de que a infraestrutura escolar é importante para o desenvolvimento dos alunos. Se não fosse, não seria alvo de políticas públicas para o seu suprimento.

Um das principais discussões sobre o tema diz respeito à quantidade de infraestrutura a ser construída. Como na ideia de utilidade marginal, é provável que o impacto no

⁶ Os trabalhos frequentemente trazem mais de uma estimação, distinguindo-as por diferentes modos de mensuração da variável dependente, por diferentes séries de ensino, ou por perfis de amostras diferentes.

⁷ Há fortes evidências de que alunos que vão bem nas disciplinas escolares ou que vão bem em testes padronizados conseguem ir mais longe na escola (Hanushek e Kimko, 2000).

desempenho dos alunos pela adição das primeiras quantidades de infraestrutura seja maior do que as adições subsequentes. Além disso, os insumos devem impactar diferentemente os alunos, principalmente se considerarmos a idade escolar deles. Por exemplo, Talvez a construção de um laboratório de química seja mais importante para estudantes de séries mais adiantadas.

Independente disso, a literatura sobre o impacto da infraestrutura é relativamente grande, e os métodos estatísticos utilizados são diversos. Contudo, dependendo do método estatístico e da amostra utilizada, o resultado pode ser diferente. Isso ocorre, para a infraestrutura, assim como para qualquer outro insumo da função de produção educacional.

Glewwe (2013) fez uma meta-análise de diversos insumos usualmente utilizados na função de produção educacional, sumarizando resultados encontrados em 79 trabalhos publicados entre 1990 e 2010. As tabelas 2.1 e 2.2 apresentam os resultados encontrados pelos artigos que tratam do tema de infraestrutura e organização escolar.

Tabela 2.1- Impacto da Infraestrutura Física da Escola na Nota em Testes

	Negativo		Zero, não significante ou	Positivo		Total de Estudos
	significativo	Não Significativo		significativo	Não Significativo	
Cadernos de Apoio/Livros Didáticos	4(3)	13(8)	7(5)	10(7)	26(10)	21
Cadeiras/Mesas/Carteiras	0(0)	0(0)	13(1)	7(5)	8(4)	8
Computadores/Jogos Eletrônicos	1(1)	9(5)	1(1)	8(3)	7(4)	8
Eletricidade	0(0)	3(2)	0(0)	6(5)	6(2)	6
Indicador de Infraestrutura da Escola	0(0)	1(1)	7(1)	1(1)	13(4)	6
Quadro Negro	0(0)	2(2)	13(1)	3(3)	7(3)	6
Biblioteca	1(1)	3(2)	7(1)	1(1)	10(5)	6
Telhado/Parede/Chão	0(0)	1(1)	0(0)	3(2)	2(1)	4

Obs: Os Valores em Parênteses Indicam a Quantidade de Estudos. Tabela Retirada de Gleww et al. (2013)

A tabela 2.1 sumariza os resultados encontrados de oito variáveis de infraestrutura e de material pedagógico. De todas essas variáveis, a literatura analisa com mais frequência o impacto de cadernos de apoio ou livros didáticos no rendimento dos alunos. Nesse caso, como mostra a tabela 2.1, foram selecionados 21 estudos contendo ao todo 60 estimativas⁸. Os números em parênteses indicam a quantidade de estudos que o resultado foi obtido. Um estudo pode ter várias estimativas diferentes, não necessariamente com o mesmo sinal, podendo mudar de acordo com a amostra ou com o método.

Apesar de não haver um consenso, a variável *Caderno de Apoio/Livros Didáticos* obteve um total de 36 resultados positivos, dos quais 26 foram significantes. Estimativas negativas e significantes foram obtidas 4 vezes em 3 estudos. Portanto,

⁸ O Somatório do número de estudos respectivo para cada variável é encontrada através da soma dos valores que estão entre parênteses na linha referente à variável, enquanto que o somatório do total de estimativas para cada variável representada na tabela é obtido pela soma dos valores fora dos parênteses na linha que a variável se encontra.

isso sugere que materiais didáticos como caderno de apoio, livros ou livros de exercícios podem aumentar as habilidades dos alunos.

Móveis e utensílios básicos para as salas de aula são a segunda variável mais analisada junto com computadores e jogos eletrônicos. O sumário indica que essas variáveis também fazem com que os alunos tenham melhor rendimento. Os 13 resultados não significativos em Cadeiras/Mesas/Carteiras foram obtidos em apenas 1 estudo.

Apenas 6 trabalhos focam na análise de um indicador de infraestrutura e apenas uma estimativa significativa foi obtida nesses trabalhos. 13 estimativas em 4 estudos obtiveram resultado positivo e não significativa e apenas um resultado negativo e não significativa. Portanto, não há um sinal conclusivo que indique que um indicador de infraestrutura gere sinal positivo.

Tabela 2.2- Impacto da Infraestrutura Organizacional da Escola na Nota em Testes

	Negativo		Zero, não significante ou	Positivo		Total de Estudos
	significativo	Não Significativo		significativo	Não Significativo	
Razão Professor/Aluno	30(13)	29(13)	3(2)	24(12)	15(9)	29
Absentismo do Professor	7(4)	6(3)	2(1)	0(0)	0(0)	5
Escola Fornece Refeições	4(1)	3(2)	0(0)	0(0)	12(3)	5
Ensino multiseriado	4(1)	0(0)	10(1)	5(2)	2(2)	4
Horas do Dia Escolar	1(1)	1(1)	0(0)	2(1)	4(2)	4

Obs: Os Valores em Parênteses Indicam a Quantidade de Estudos. Tabela Retirada de Gleww et al. (2013)

A tabela 2.2 analisa a estrutura organizacional da escola. Largamente, a variável razão professor/aluno é a mais estudada em artigos da área. Entretanto, não há qualquer sinal de que haja uma influência positiva ou negativa do tamanho da turma no desempenho dos alunos.

O fato de a escola fornecer refeições e dos professores se absterem com frequência parece gerar um efeito negativo sobre os alunos. Enquanto que ensino multisseriado e a quantidade de horas-aula dos alunos parecem ter um efeito positivo. Mas a quantidade de estudos é insuficiente para dar um possível sinal da direção dos efeitos.

3. Metodologia e Base de Dados

Como já explicitado, o foco principal deste trabalho é entender como a infraestrutura escolar, considerada tanto em termos físicos como administrativos, afeta, se ela de fato afeta, o desempenho dos alunos em testes padronizados de ensino. Nossa primeira hipótese é que a infraestrutura escolar afeta positivamente o rendimento dos alunos. Supomos ainda que esse efeito é heterógeno ao longo da distribuição de conhecimento dos alunos. Para isso, utilizaremos os dados do Sistema de Avaliação da Educação Básica (Saeb) do ano de 2011, em que consta além da nota dos alunos no exame, questionário respondido pelos alunos sobre sua situação na família e de características pessoais ligadas a ele próprio, questionário respondido pelo professor e outro pelo diretor ambos sobre características profissionais e da escola e um questionário respondido pelo entrevistador sobre a estrutura física da escola. A estratégia de estimação se baseia por Mínimos Quadrados Ordinários e por Regressão Quantílica.

Neste capítulo, serão explicitados na primeira seção como se configura o Saeb a composição da base de dados, a fonte, o modelo econométrico e o tratamento aos dados. Na segunda seção serão abordados os modelos econométricos utilizados.

3.1. Base de Dados

O Saeb tem como principal objetivo avaliar a Educação Básica brasileira e contribuir para a melhoria de sua qualidade e para a universalização do acesso à escola, oferecendo subsídios concretos para a formulação, reformulação e o monitoramento das políticas públicas voltadas para a Educação Básica. Além disso, procura também oferecer dados e indicadores que possibilitem maior compreensão dos fatores que influenciam o desempenho dos alunos nas áreas e anos avaliados⁹.

O Saeb é composto por três avaliações externas em larga escala:

Avaliação Nacional da Educação Básica – Aneb: abrange, de maneira amostral, alunos das redes públicas e privadas do país, em áreas urbanas e rurais, matriculados na 4ª série/5ºano e 8ªsérie/9ºano do Ensino Fundamental e no 3º ano do Ensino Médio, tendo como principal objetivo avaliar a qualidade, a equidade e a eficiência da educação brasileira. Apresenta os resultados do país como um todo.

Avaliação Nacional do Rendimento Escolar - Anresc (também denominada "Prova Brasil"): trata-se de uma avaliação censitária envolvendo os alunos da 4ª série/5ºano e 8ªsérie/9ºano do Ensino Fundamental das escolas públicas das redes municipais,

⁹ Essas informações foram retiradas do site: <http://portal.inep.gov.br/web/saeb/aneb-e-anresc>

estaduais e federal, com o objetivo de avaliar a qualidade do ensino ministrado nas escolas públicas. Participam desta avaliação as escolas que possuem, no mínimo, 20 alunos matriculados nas séries/anos avaliados, sendo os resultados disponibilizados por escola, por município e por estado.

A Avaliação Nacional da Alfabetização – ANA : avaliação censitária envolvendo os alunos do 3º ano do Ensino Fundamental das escolas públicas, com o objetivo principal de avaliar os níveis de alfabetização e letramento em Língua Portuguesa, alfabetização Matemática e condições de oferta do Ciclo de Alfabetização das redes públicas. A ANA foi incorporada ao Saeb pela Portaria nº 482, de 7 de junho de 2013.

A edição com disponibilidade de dados disponível mais recentemente é a edição de 2011, que é a base utilizada neste trabalho. As escolas avaliadas pela Aneb e Anresc (Prova Brasil) dividem-se em dois grupos: as que foram avaliadas de forma censitária e as que foram avaliadas de modo amostral, por sorteio. Na edição de 2011, 55.924 escolas públicas participaram da parte censitária, a chamada Prova Brasil, e 3.392 escolas públicas e particulares participaram da parte amostral.

O grupo da prova Brasil de escolas recebeu aplicação censitária em turmas de 5º e 9º anos do ensino fundamental público, nas redes estaduais, municipais e federais, de área rural e urbana, desde que a escola possuísse no mínimo 20 alunos matriculados em cada série avaliada. Para esse grupo, os resultados são divulgados por escola.

Já a parte amostral da avaliação abrangeu escolas com 10 a 19 alunos de 5º e 9º anos do ensino fundamental das redes públicas; escolas com 10 ou mais alunos de 5º e 9º anos do ensino fundamental das redes privadas; e escolas com 10 ou mais alunos da 3ª série do ensino médio das redes públicas e privadas do país.

A metodologia da Aneb e Anresc (Prova Brasil) baseia-se na aplicação de testes padronizados de Língua Portuguesa e Matemática e de Questionários Socioeconômicos a estudantes de 5º ano e 9º ano do Ensino Fundamental e 3ª série do Ensino Médio. Além dos estudantes, diretores e professores também preenchem ao Questionário Socioeconômico.

O plano amostral da Aneb e Anresc (Prova Brasil) permite que os resultados sejam apresentados para cada município, Unidade da Federação, Região e para o Brasil como um todo, por dependência administrativa, localização e área em cada uma das agregações possíveis.

A base metodológica das duas provas é a mesma, o que difere uma prova da outra é a população de estudantes a qual o exame é aplicado. A Aneb é avaliada por amostra e é possível que nem todas as turmas e nem todos os estudantes de uma escola sorteada sejam avaliados.

Ao todo, são confeccionados 21 tipos diferentes de cadernos de prova para cada série, sendo que cada aluno responde a apenas um caderno de prova. Desta forma, dois alunos não respondem necessariamente às mesmas questões.

Cada caderno de prova é constituído por quatro blocos, sendo que dois são destinados a respostas de Língua Portuguesa e os outros dois abordam questões de Matemática. Os testes são de múltipla escolha, com quatro ou cinco alternativas de resposta para cada questão, sendo que apenas uma está correta.

Existem, no total, 77 itens de cada disciplina na 4ª série e 91 itens de cada disciplina na 8ª série do Ensino Fundamental e no 3º ano do Ensino Médio distribuídos pelos 21 cadernos de prova.

As notas são calculadas de uma forma as médias são apresentadas em uma escala capaz de descrever, em cada nível, as competências e as habilidades que os estudantes demonstram ter desenvolvidos. Há uma escala contínua descrita para as habilidades em língua portuguesa e matemática.

A escala das notas é única e cumulativa para todas as séries avaliadas. Dessa forma, é comum que os estudantes da quinta série apresentem notas inferiores aos alunos do nono ano e assim sucessivamente.

Para este estudo a base de dados se restringirá apenas a alunos de escolas públicas para todas as séries. Isso se deve ao foco do estudo em políticas públicas e para minimizar possíveis vieses ocasionados pelo fato de estudantes de escolas particulares, além de possuírem de modo geral acesso a melhor infraestrutura escolar, possuem melhores condições no suporte familiar. Em escolas públicas, a entrada é livre, portanto, o viés de seleção, a princípio não existe. É claro que a localidade da escola possui alguma interferência na seleção dos alunos, no entanto, supomos que não haja barreiras à entrada.

Na tabela 3.1 encontram-se as proficiências de matemática para alguns pontos de distribuição. Observa-se que as notas médias são semelhantes às medianas. Além disso, as proficiências são maiores nos percentis superiores. Nas tabelas 3.2, 3.3 e 3.4 encontram-se algumas características dos alunos de acordo com a posição da distribuição de proficiência. É possível observar nos percentis superiores de notas uma maior proporção de alunos com melhor *background* familiar.

Tabela 3.1- Distribuição de Proficiência de Matemática

	Média	Desvio	10°	20°	30°	40°	Mediana	60°	70°	80°	90°
Quinta Série	208,87	47,52	148,44	164,46	177,4	189,81	202,75	216,3	231,22	248,63	273,34
Nona Série	247,48	46,37	183,66	204,26	212,3	244,79	257,22	270,33	277,53	285,35	305,25
Terceiro Ano	292,14	63,54	203,8	231,81	252,94	271,77	290,4	308,55	328,23	351,56	380,62

Fonte: Elaboração Própria Utilizando Microdados do Saeb 2011

Tabela 3.2- Características dos Alunos Segundo a Distribuição das Proficiências de Matemática-Quinta Série

	Média (%)	Desvio (%)	10° Decil (%)	20° Decil (%)	30° Decil (%)	40° Decil (%)	50° Decil (%)	60° Decil (%)	70° Decil (%)	80° Decil (%)	90° Decil (%)
Branco	,30	,46	,26	,21	,23	,27	,28	,29	,30	,37	,37
Tem computador com internet em casa	,39	,49	,25	,28	,33	,34	,37	,38	,46	,46	,57
Mora com 4 ou mais pessoas	,65	,48	,82	,73	,68	,70	,68	,64	,59	,55	,50
Lê livros	,41	,49	,37	,38	,38	,37	,40	,41	,39	,41	,37
Frequenta bibliotecas	,24	,43	,32	,24	,27	,22	,25	,22	,22	,28	,17
Trabalha em casa 2 horas ou mais	,56	,50	,70	,61	,61	,59	,59	,54	,47	,50	,55
Trabalha fora de casa	,14	,34	,19	,20	,13	,13	,13	,10	,10	,08	,06
Já foi reprovado	,31	,46	,52	,51	,44	,37	,31	,31	,23	,19	,13
Faz dever de matemática	,75	,43	,66	,65	,69	,72	,72	,74	,82	,83	,84
Creche	,37	,48	,34	,38	,36	,36	,36	,34	,35	,38	,34
Pré-escola	,37	,48	,33	,29	,34	,32	,35	,39	,41	,40	,47
Pai ou mãe tem ensino superior	,15	,36	,13	,12	,15	,15	,15	,11	,16	,14	,17
Infraestrutura	,07	,41	-,07	-,04	,00	,06	,06	,10	,13	,14	,15

Fonte: Elaboração Própria Utilizando Microdados do Saeb 2011

Tabela 3.3- Características dos Alunos Segundo a Distribuição das Proficiências de Matemática-Nona Série

	Média (%)	Desvio (%)	10° Decil (%)	20° Decil (%)	30° Decil (%)	40° Decil (%)	50° Decil (%)	60° Decil (%)	70° Decil (%)	80° Decil (%)	90° Decil (%)
Tem computador com internet em casa	,49	,50	,34	,35	,36	,45	,50	,53	,51	,55	,54
Mora com 4 ou mais pessoas	,54	,50	,64	,62	,63	,53	,50	,50	,50	,48	,39
Lê livros	,30	,46	,29	,30	,27	,30	,24	,29	,28	,27	,39
Frequenta bibliotecas	,13	,34	,18	,13	,17	,13	,11	,13	,12	,11	,13
Trabalha em casa 2 horas ou mais	,59	,49	,61	,60	,64	,62	,59	,62	,56	,59	,53
Trabalha fora de casa	,19	,39	,23	,19	,21	,20	,19	,17	,18	,16	,14
Já foi reprovado	,32	,47	,50	,45	,41	,33	,33	,30	,26	,24	,19
Faz dever de matemática	,57	,49	,46	,48	,52	,53	,53	,61	,60	,68	,65
Branco	,34	,47	,29	,28	,27	,30	,33	,33	,36	,43	,42
Creche	,32	,47	,30	,33	,38	,34	,33	,31	,29	,32	,38
Pré-escola	,47	,50	,43	,41	,38	,46	,48	,49	,52	,51	,49
Infraestrutura	,09	,34	,01	,05	,05	,10	,09	,09	,12	,11	,14
Pai ou mãe tem ensino superior	,11	,31	,05	,06	,06	,10	,11	,11	,10	,16	,12

Fonte: Elaboração Própria Utilizando Microdados do Saeb 2011

Tabela 3.4- Características dos Alunos Segundo a Distribuição das Proficiências de Matemática-Terceiro Ano

	Média (%)	Desvio (%)	10° Decil (%)	20° Decil (%)	30° Decil (%)	40° Decil (%)	50° Decil (%)	60° Decil (%)	70° Decil (%)	80° Decil (%)	90° Decil (%)
Tem computador com internet em casa	,61	,49	,38	,34	,53	,50	,59	,65	,77	,82	,92
Lê livros	,49	,50	,54	,48	,54	,43	,44	,45	,40	,45	,44
Trabalha em casa 2 horas ou mais	,52	,50	,57	,67	,55	,67	,56	,56	,45	,35	,27
Trabalha fora de casa	,25	,44	,37	,35	,42	,29	,30	,28	,26	,24	,13
Já foi reprovado	,25	,43	,54	,42	,32	,29	,24	,24	,24	,12	,08
Faz dever de matemática	,55	,50	,47	,52	,48	,55	,51	,50	,52	,50	,57
Branco	,39	,49	,24	,24	,35	,28	,39	,38	,47	,54	,54
Creche	,31	,46	,27	,29	,30	,32	,28	,25	,31	,35	,47
Pré-escola	,51	,50	,44	,45	,52	,47	,56	,55	,57	,59	,47
Infraestrutura	,32	,70	-,04	,02	-,04	,21	,20	,34	,46	,66	,78
Pai ou mãe tem ensino superior	,28	,45	,10	,12	,07	,19	,21	,27	,37	,40	,57
Mora com 4 ou mais pessoas	,44	,50	,50	,51	,38	,44	,43	,42	,34	,31	,27

Fonte: Elaboração Própria Utilizando Microdados do Saeb 2011

O primeiro desafio é escolher uma variável que sinalize se a escola possui ou não infraestrutura adequada para acolher os estudantes. Boa parte dos estudos com função de produção educacional foca em analisar como se comporta o desempenho dos alunos sob alguns componentes que podem indicar o grau de estrutura da escola.

Como comentado em Gleww et al. (2013), os estudos ligados à indicadores de infraestrutura física em geral destacam o estudo de variáveis como, “a escola possui carteiras, quadros, mesas e cadeiras em bom estado”, “a fiação elétrica e sistema de água e esgoto disponível ou em bom estado” e “a escola possui biblioteca ou não”. Do ponto de vista da estrutura organizacional, os estudos focam na razão professor/aluno em sala de aula, no salário do professor, no valor despendido por aluno e na quantidade de horas que o aluno passa dentro da escola fazendo alguma atividade. Eventualmente, há trabalhos que montam indicadores de qualidade da escola ou indicadores de infraestrutura. Essas variáveis recebem controle de outras que incorporam características dos estudantes ou de *background* familiar.

Tendo a quantidade enorme de indicadores de infraestrutura que os questionários do Saeb fornecem, é interessante buscar a fabricação de um indicador de infraestrutura que comporte todas as variáveis de infraestrutura de interesse para fins desse trabalho.

As tabelas 3.2, 3.3 e 3.4 destacam a variável infraestrutura, que será utilizado nas regressões para estimar a influência delas nas notas dos alunos no teste do Saeb. Essa variável é formada por um indicador composto de inúmeras outras variáveis que indicam o nível de infraestrutura física e administrativa das escolas.

Para montar esse indicador, foram selecionadas as variáveis descritas no Quadro 3.1, composto em sua maioria por *dummies*. Após essa seleção, fez-se a padronização dessas variáveis e as regressiu contra a nota padronizada da escola (nota média dos alunos da série de interesse). Serão feitas três regressões, uma para cada série de ensino. As únicas variáveis que diferem de uma regressão para a outra é a nota padronizada da escola, como variável dependente.

Algumas vezes é útil obter resultados de regressão quando todas as variáveis envolvidas, a dependente e todas as independentes, tenham sido padronizadas. Uma variável é padronizada em uma amostra pela subtração de sua média e dividindo o resultado por seu desvio-padrão. Isso significa que computamos a transformação de cada variável na amostra. Depois, fazemos a regressão usando os valores da “nova variável”.

Começemos com a equação MQO original, com suas variáveis em suas formas originais:

$$y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \widehat{\beta}_2 x_{i2} + \dots + \widehat{\beta}_3 x_{ik} + \widehat{u}_i \quad (3.1)$$

Incluimos o subscrito de observação i para enfatizar que nossa padronização é aplicada a todos os valores da amostra. Agora, se ao calcularmos a média de (3.1), usarmos o fato de que \widehat{u}_i tem média de amostra zero, e subtrairmos o resultado de (3.1), temos:

$$y_i - \bar{y} = \widehat{\beta}_1 (x_{i1} - \bar{x}_1) + \widehat{\beta}_2 (x_{i2} - \bar{x}_2) + \dots + \widehat{\beta}_3 (x_{ik} - \bar{x}_k) + \widehat{u}_i \quad (3.2)$$

Em seguida, definimos $\widehat{\sigma}_y$ como o desvio-padrão da amostra da variável dependente, $\widehat{\sigma}_1$ como o desvio padrão da amostra da x_1 , $\widehat{\sigma}_2$ como desvio padrão de x_2 , e assim sucessivamente. Agora, segue a equação:

$$\left(\frac{y_i - \bar{y}_i}{\widehat{\sigma}_y} \right) = \left(\frac{\widehat{\sigma}_1}{\widehat{\sigma}_y} \right) \widehat{\beta}_1 \left(\frac{x_i - \bar{x}_i}{\widehat{\sigma}_1} \right) + \dots + \left(\frac{\widehat{\sigma}_n}{\widehat{\sigma}_y} \right) \widehat{\beta}_n \left(\frac{x_n - \bar{x}_n}{\widehat{\sigma}_n} \right) + (\widehat{u}_i / \widehat{\sigma}_y) \quad (3.3)$$

Cada variável em (3.3) foi padronizada e isso resultou em novos coeficientes de inclinação. O intercepto desaparece e o coeficiente de inclinação de $\frac{x_i - \bar{x}_i}{\widehat{\sigma}_1}$ é $\left(\frac{\widehat{\sigma}_1}{\widehat{\sigma}_y} \right) \widehat{\beta}_1$, por exemplo.

Os novos coeficientes representados na equação (3.3) significam que se x_1 aumentar em um desvio padrão, então \widehat{y} será alterado em $\left(\frac{\widehat{\sigma}_1}{\widehat{\sigma}_y} \right) \widehat{\beta}_1$ desvios padrão. Com isso, estamos medindo os efeitos não em termos das unidades originais de y ou de x_j , mas em unidades de desvio padrão. A dimensão dos regressores, portanto, seria

irrelevante. Essas variáveis são colocadas por essa equação com o mesmo poder de influência.

As estimações desses coeficientes padronizados estão no apêndice 1. É interessante notar que o R^2 é pequeno para todas as séries de ensino, entretanto, ele é menor para o terceiro ano, podendo significar que a infraestrutura escolar é menos importante para as séries finais. Essa é uma possibilidade que é fortalecida pela presença apenas de 14 variáveis significantes a 5% de 21 ao todo para o terceiro ano do ensino médio, enquanto que apenas 2 variáveis na quinta e na nona série são não significantes a 5%.

O indicador é finalmente montado pela multiplicação entre os coeficientes do Apêndice 1 que são significantes a 5% e o valor das próprias variáveis. Então se as variáveis são dummies, o coeficiente das regressões do Apêndice 1 são agregadas ou não ao indicador.

Utilizar variáveis relacionadas à infraestrutura em modelos de função de produção educacional não é algo simples. Normalmente, os sinais das variáveis alteram muito dependendo do tipo de aluno, do tipo de escola e do tipo de amostra. No nosso caso não foi diferente, as variáveis trocam muito de sinal ou até mudam de intensidade de acordo com a série de ensino, sendo que muitas vezes não há uma lógica clara porque acontecem essas mudanças. É difícil explicar como o fato de não ocorrer falta de professores naquele ano ter impacto positivo na nota média das notas na quinta série, ter impacto negativo na nona série e voltar a ter impacto positivo no terceiro ano.

As variáveis que claramente apresentam sinal positivo e significativa a pelo menos 1% para as séries em geral são: “computadores para os alunos”, “computadores para uso administrativo”, “máquina copiadora”, “biblioteca” e “quadra de esportes”, apesar de “computadores para os alunos” não ser significativo no terceiro ano. Por outro lado, as variáveis que apresentam sinal negativo mais claro são “sinais de depredação”, “mesas para cadeirantes”, “atividades artísticas” e “acessibilidade”, apesar de para a nona série o sinal de a primeira ser ambíguo e não significativo, e da segunda, da terceira e da quarta os sinais serem ambíguos e não significantes para o terceiro ano.

O fato é que o objetivo deste trabalho não é afirmar quais variáveis devem ser utilizadas como políticas públicas ou as que não devem e sim, chamar a atenção para a importância da infraestrutura como um todo. Os sinais descrevem o impacto médio da média da nota dos alunos para cada escola avaliada. É óbvio que não seria desejoso que coeficientes tivessem tantas variações de sinal, mas se a variável apresenta na média impacto positivo para a nota da escola, foi porque na média a variável também mostrou impacto positivo para os estudantes.

Quadro 3.1- Descrição das Variáveis Utilizadas para a Montagem do Indicador de Infraestrutura

Variáveis	Descrição
Sala de aula em bom estado	<i>Padronização da dummy igual a 1 se o estado de conservação das salas de aula se encontra em bom estado, 0 caso contrário.</i>
Banheiros em bom estado	<i>Padronização da dummy igual a 1 se o estado de conservação dos banheiros se encontra em bom estado, 0 caso contrário.</i>
Instalação elétrica em bom estado	<i>da instalação elétrica se encontra em bom estado, 0 caso contrário.</i>
Sinais de depredação	<i>Padronização da dummy igual a 1 se a escola apresenta sinais de depredação (vidros, portas e janelas quebradas, lâmpadas estouradas etc.), 0 caso contrário.</i>
Computadores para os alunos	<i>Padronização da dummy igual a 1 se a escola possui computadores com acesso aos alunos, 0 caso contrário.</i>
Computadores para uso administrativo	<i>Padronização da dummy igual a 1 se a escola possui computadores para uso administrativo, 0 caso contrário.</i>
Máquina Copiadora	<i>Padronização da dummy igual a 1 se a escola possui máquina copiadora, 0 caso contrário.</i>
Biblioteca	<i>Padronização da dummy igual a 1 se a escola possui biblioteca, 0 caso contrário.</i>
Quadra de esportes	<i>Padronização da dummy igual a 1 se a escola possui quadra de esportes, 0 caso contrário.</i>
laboratório	<i>Padronização da dummy igual a 1 se a escola possui biblioteca, 0 caso contrário.</i>
Sala de música	<i>Padronização da dummy igual a 1 se a escola possui sala de música, 0 caso contrário.</i>
Sala de artes	<i>Padronização da dummy igual a 1 se a escola possui sala de artes, 0 caso contrário.</i>
Sala de leitura	<i>Padronização da dummy igual a 1 se a escola possui sala de leitura, 0 caso contrário.</i>
Programa de redução de taxa de reprovação	<i>Padronização da dummy igual a 1 se a escola possui Programa de redução de taxa de reprovação, 0 caso contrário.</i>
Programa de reforço escolar	<i>Padronização da dummy igual a 1 se a escola possui Programa de reforço escolar, 0 caso contrário.</i>
Não ocorreu problemas financeiros	<i>pesquisa não houve problema de caráter financeiro, 0 caso contrário.</i>
Não ocorreu falta de professor	<i>Padronização da dummy igual a 1 se na escola no ano da pesquisa não houve falta de professor, 0 caso contrário.</i>
Acessibilidade	<i>Padronização da dummy igual a 1 se no interior da escola, na entrada principal e nas circulações internas: há faixa de circulação livre de obstáculos com largura mínima de 80 cm</i>
Mesas para cadeirantes	<i>Padronização da dummy igual a 1 se na escola possui: mesas e cadeiras acessíveis para pessoas em cadeiras de rodas, 0 caso</i>
Atividades esportivas	<i>Padronização da dummy igual a 1 se na escola possui regularmente atividades esportivas, 0 caso contrário.</i>
Atividades artísticas	<i>Padronização da dummy igual a 1 se na escola possui regularmente atividades artísticas, 0 caso contrário.</i>

Fonte: Elaboração Própria Utilizando Microdados do Saeb 2011

Depois de feita a regressão, o indicador é formado pela soma dos coeficientes das variáveis selecionadas que obtiveram significância de pelo menos 5% multiplicadas pela própria variável padronizada (as regressões, com os coeficientes das variáveis relacionadas à infraestrutura, estão em anexo no Apêndice 1)¹⁰.

Dessa forma, é formado um indicador de infraestrutura para cada série de cada escola diferente. A estrutura dele é contínua e vai de -1,53 até 1,50 para as escolas com referência a quinta séries das escolas, de -1,28 até 2,88 para as escolas com referência à oitava série e -0,85 até 1,76 para as escolas com referência ao terceiro ano.

Por fim, no quadro 3.2 encontram-se as variáveis que foram usadas nas estimações. Fez-se controle principalmente por variáveis que representam o *background* familiar, mas também se tentou captar o efeito do esforço do aluno através das variáveis “lê livros” e “faz dever de matemática”. Além disso, controlou-se pela região onde o estudante mora, tendo como base a região Sudeste do país.

Para as regressões, foram tomadas amostras aleatórias da base de dados do Saeb, embora haja acesso a uma base praticamente censitária. Para a quinta série foi pego uma amostra de 92.692 alunos, de 79.631 alunos para a nona série e de 68.995 alunos para o terceiro ano. A razão disso é a dificuldade em processar por regressão quantílica a grande quantidade de dados. No entanto, de acordo com a regressão em MQO, não houve grandes mudanças nos estimadores, nem na significância deles. O mesmo deve acontecer com respeito à regressão quantílica. Ambos os modelos serão revisados na próxima seção.

¹⁰ Esse indicador de infraestrutura tem como objetivo selecionar um grande número de variáveis relacionadas à infraestrutura física e organizacional da escola, considerando questões sobre as condições das salas de aula, acessibilidade, disponibilidade de recursos e organização dos alunos nas salas de aula. A partir disso o indicador é ponderado de acordo com a importância de cada uma dessas variáveis na nota das escolas, esta obtida a partir da nota média dos alunos para cada série. Portanto, se por acaso a biblioteca tem um coeficiente positivamente grande na regressão entre as notas padronizadas das escolas e as variáveis de infraestrutura padronizadas e um estudante não está inserido numa escola que possui biblioteca, então o aluno terá um indicador de infraestrutura bem menor que os demais.

Quadro 3.2- Descrição das Variáveis Utilizadas no Modelo

Variáveis	Descrição
Infraestrutura	<i>Indicador de Infraestrutura representativo da escola para aquela série ao qual o aluno pertence.</i>
Branco	<i>Dummy igual a 1 se o aluno é branco, 0 caso contrário.</i>
Tem computador com internet em casa	<i>Dummy igual a 1 se o aluno possui em casa computador com acesso a internet, 0 caso contrário.</i>
Mora com 4 ou mais pessoas	<i>Dummy igual a 1 se o aluno mora com 4 pessoas ou mais, 0 caso contrário.</i>
Lê livros	<i>Dummy igual a 1 se o aluno lê livros sempre, 0 caso contrário.</i>
Frequenta bibliotecas	<i>Dummy igual a 1 se o aluno frequenta bibliotecas sempre, 0 caso contrário.</i>
Faz Trabalhos Domésticos	<i>Dummy igual a 1 se o aluno faz trabalhos domésticos ao menos 2 horas por dia, 0 caso contrário.</i>
Trabalha fora de casa	<i>Dummy igual a 1 se o aluno trabalha fora de casa, 0 caso contrário.</i>
Já foi reprovado	<i>Dummy igual a 1 se o aluno já foi reprovado pelo menos uma vez, 0 caso contrário.</i>
Faz dever de matemática	<i>Dummy igual a 1 se o aluno faz sempre trabalho doméstico, 0 caso contrário.</i>
Creche	<i>Dummy igual a 1 se o aluno iniciou sua vida escolar na creche, 0 caso contrário.</i>
Pré-escola	<i>Dummy igual a 1 se o aluno iniciou sua vida escolar na pré-escola, 0 caso contrário.</i>
Pai ou mãe tem ensino superior	<i>Dummy igual a 1 se o aluno possui pai ou mãe com ensino superior, 0 caso contrário.</i>
Norte	<i>Dummy igual a 1 se o aluno nasceu na região Norte, 0 caso contrário.</i>
Nordeste	<i>Dummy igual a 1 se o aluno nasceu na região Nordeste, 0 caso contrário.</i>
Sul	<i>Dummy igual a 1 se o aluno nasceu na região Sul, 0 caso contrário.</i>
Centro Oeste	<i>Dummy igual a 1 se o aluno nasceu na região Centro-Oeste, 0 caso contrário.</i>

Fonte: Elaboração Própria Utilizando Microdados do Saeb 2011

3.2 Metodologia

A estimação do parâmetro de infraestrutura será feito por meio do método de Mínimos Quadrados Ordinários (MQO) e também pelo método de Regressão Quantílica. Ambos os métodos fazem controle pelas características de background familiar dos alunos, como educação dos pais, de características pessoais, como hábito de leitura, e da região onde estuda (entendendo-se como a região geográfica no país).

Certamente, boa parte dos trabalhos econométricos aplicados se preocupam basicamente com a média. O foco em médias é parcialmente causado porque a obtenção de uma boa estimativa da causalidade do efeito médio é forte o suficiente. Se a variável dependente for uma dummy, a média descreve a distribuição inteira. Entretanto, muitas variáveis, como a nota em testes de proficiência, são variáveis

continuamente distribuídas. Essas distribuições podem não ser reveladas completamente com a análise da média. Por exemplo, a distribuição pode alternar partes mais espaçadas ou mais condensadas dentro da população ou amostra. Por conta disso, é importante buscar saber o que está havendo dentro de toda a distribuição.

O que a curva de regressão faz é dar um grande resumo das médias das distribuições correspondentes ao conjunto das variáveis independentes observadas. Poderíamos buscar mais informações computando diversas curvas de regressão correspondendo aos vários pontos percentuais da distribuição e dessa forma ter uma visão mais completa do conjunto de distribuição.

Por conta disso, a regressão em MQO não será suficiente para nosso estudo. Será necessário utilizar o ferramental da regressão quantílica.

Na primeira subseção será abordado um pequeno resumo de como a Função de produção educacional é tratada sob a ótica dos Mínimos Quadrados Ordinários, explicitando sua funcionalidade e o problema de endogeneidade. Na segunda subseção será introduzida a técnica de regressão quantílica linear, que resolve o problema de uma curva de regressão única para a média da distribuição.

3.2.1 A Abordagem da Função de Produção Educacional no Modelo de Mínimos Quadrados Ordinários (MQO)

Para fins deste trabalho faz-se uso de uma função de produção educacional em que a variável dependente é a nota do aluno em matemática no teste padronizado obtido pelo Saeb. As variáveis explicativas são os insumos oferecidos pela escola, as características da família, da residência, da escola¹¹. Deste modo, podemos formalizar essa relação da seguinte maneira:

$$y_{(i)} = \beta_0 + \beta_{fam}X_{fam_i} + \beta_{esc}X_{esc_i} + \beta_{aluno}X_{aluno_i} + \beta_{região}X_{região_i} + u_{(i)}$$

Em que a nota no teste de proficiência é individualizada e representada pelo subscrito $i = 1, 2, \dots, N$. Em que X_{fam} é vetor de variáveis com respeito a características da família, X_{esc} é o vetor de variáveis com respeito a características da escola e X_{aluno} é o vetor de variáveis com respeito a características do próprio aluno e $X_{região}$ é uma dummy que representa a região geográfica em que o aluno estuda.

A equação linear em (3.4) é inspirada no modelo representativo em (2.1). Como o foco deste trabalho é estudar o efeito da infraestrutura nas escolas no rendimento dos

¹¹ Mais adiante será exibido detalhamento da fonte dos dados e como eles foram tratados para que o modelo fosse rodado.

alunos, foi retirada do modelo características com respeito ao professor. Adicionou-se também um dummy com respeito à região geográfica de onde fica a escola para tentar minimizar algum viés de variável omitida, lembrando que o estudo abrange todas as regiões do Brasil. Além disso, supõe-se que os pais só tem apenas uma opção de escola para a matrícula dos filhos. Essa hipótese afirma que a característica escolar é exógeno ao perfil da família, o que minimiza o viés de simultaneidade.

A justificativa para utilizar o MQO em vez de outros estimadores concorrentes é dada pelo teorema de Gauss-Markov. Para isso é necessário que algumas hipóteses sejam explicitadas. A primeira hipótese do modelo de regressão linear múltipla (RLM.1) é a linearidade nos parâmetros. A segunda é a amostragem aleatória (RLM.2). A terceira hipótese e, a mais implicante delas, é a de média condicional zero (RLM.3) dos erros do modelo. Formalmente:

$$E(u|x_1, x_2, \dots, x_k)=0$$

Esta hipótese implica que as variáveis não incluídas no modelo, ou o erro, sejam não correlacionados com as variáveis incluídas. A violação dessa hipótese pode ocorrer basicamente quando o modelo está mal especificado ou quando há omissão de variável relevante. O terceiro problema é geralmente mais sensível que os dois primeiros e pode afetar particularmente estudos empíricos envolvendo características populacionais. No contexto do trabalho imagine que a nota do aluno seja influenciada pela didática do professor. No entanto, é praticamente impossível obter uma fonte de dados em que seja demonstrada a aptidão do professor para uma boa didática no ensino. É bem provável que a didática do professor seja relacionada com o salário que ele recebe e com a característica dos alunos, como atenção ao professor, dedicação em sala de aula, etc.

O problema é que o impacto causal da variável observada, seja ela a de infraestrutura, no aprendizado só pode ser consistentemente estimado por mínimos quadrados ordinários se o termo erro for não correlacionada com todas as outras variáveis no modelo, o que é praticamente impossível.

As maiores preocupações com a modelagem da função de produção são o viés de variável omitida, seleção da amostra, simultaneidade e erro de mensuração. No primeiro problema, se quaisquer variáveis que impactam o alcance dos estudantes forem omitidas da estimação da equação e se essas variáveis forem correlacionadas com as variáveis que entraram no modelo, então o viés é introduzido, sendo o viés proporcional à importância da variável omitida. Fatores ligados aos professores e a escola influenciam na escolha por parte dos pais em que escola o filho irá estudar. A qualidade da escola também pode afetar a entrada de mais recursos governamentais na escola. O primeiro caso subestima o poder de influência dos professores e da escola, já o segundo superestima a influência dos recursos financeiros.

Estudos mais recentes focam cada vez mais no uso de ensaios controlados aleatórios e em dados em painel para fazer as estimações controlando toda a quantidade de variáveis possíveis.

Assume-se que os problemas tratados com relação aos problemas de omissão de variável são importantes e são reais nesse trabalho. Entretanto, o foco aqui é o desenvolvimento de um indicador de infraestrutura que inclua a maior quantidade possível de variáveis ligadas a estrutura de uma escola e estimá-lo contra a nota dos alunos e saber como ele afeta o resultado dos estudantes na média e ao longo da distribuição.

A quarta hipótese é a colinearidade não perfeita (RLM.4). Esta hipótese diz apenas que as variáveis independentes não podem ser constantes e que não pode haver relações exatas entre as variáveis independentes. Sob estas hipóteses citadas os parâmetros estimados serão não-viesados dos parâmetros da população.

É importante observar um ponto: o problema da multicolinearidade. Embora esse problema não viole a hipótese de inexistência de viés, a multicolinearidade, causada pela elevada correlação entre duas ou mais variáveis independentes gera um alto valor para as variâncias amostrais dos estimadores de inclinação de MQO, este por sua vez causa menor imprecisão aos estimadores de MQO.

A fim de eliminar a suspeita de haver multicolinearidade, será realizado um teste de Fatores de Inflação da Variância (VIF), que formalmente é representado por:

$$VIF = \frac{1}{(1-R_k^2)}$$

Em que R_k^2 é o r-quadrado da k-ésima variável independente obtido pela regressão da k-ésima variável independente pelas outras variáveis independentes. Se $VIF > 10$, então há presença de multicolinearidade.

Por fim, outra grande preocupação nos modelos experimentais é o problema da heterocedasticidade. Esse problema viola a hipótese de que a variância do termo erro, u , condicionada às variáveis explicativas, é a mesma para as combinações de resultados das variáveis explicativas, formalmente:

$$Var(u|x_1, \dots, x_k) = 0$$

A homocedasticidade não se mantém sempre que a variância dos fatores não observáveis muda ao longo dos estratos populacionais, nos quais esses mesmos estratos são determinados pelas variáveis explicativas. A heterocedasticidade é algo bastante comum em regressões, principalmente quando se deseja extrair informações sobre determinados grupos populacionais.

No presente estudo a heterocedasticidade é algo talvez menos preocupante do que a omissão de variável, mas deve ser um problema tratado, ainda mais porque nada garante como um determinado grupo de alunos reage sobre alguma variável disponível.

Por um lado, a heterocedasticidade não provoca viés ou inconsistência nas estimações dos coeficientes, como a omissão de uma variável importante, no entanto, ela se justifica pela aplicação imperfeita da análise econométrica e da inferência habitual por MQO. Sem a hipótese de homocedasticidade, os estimadores das variâncias, $Var(\tilde{\beta}_i)$, são viesados. Por sua vez, os erros-padrão, utilizados para a formação da estatística t e dos intervalos de confiança, são construídos através dessas variâncias. Como a variância não é constante, as estatísticas t habituais dos estimadores de MQO não apresentam distribuição t. O mesmo vale para as estatísticas F e LM, que não irão apresentar distribuição F e qui-quadrado assintótica, respectivamente.

Entretanto, o ajuste dos erros padrões por meio de procedimento *robusto em relação à heterocedasticidade* torna válida a inferência e as estatísticas dos estimadores na presença de heterocedasticidade de forma desconhecida. Podemos testar também a presença da heterocedasticidade por meio do teste Breusch-Pagan, em que um p-valor suficientemente pequeno caracteriza a presença do problema.

Embora o método MQO seja menos eficaz para tratar com heterocedasticidade, os modelos de regressão quantílica conseguem incorporar e minimizar esse problema. A propósito, uma evidência de que há ocorrência de heterocedasticidade é quando na regressão quantílica os coeficientes das variáveis ao longo dos quantis de distribuição possuem valores muito diferentes.

Não é sempre que a hipótese de normalidade na distribuição do erro é concretizada. Porém, mesmo que não se observe distribuição normal dos erros, é possível utilizar o teorema do limite central para concluir que os estimadores de MQO satisfazem a normalidade assintótica. Isto é, em amostras grandes a distribuição t se aproxima cada vez mais da distribuição da estatística t verdadeira, assim como a estatística F se aproxima da distribuição da estatística F verdadeira.

3.1.2 A Abordagem da Função de Produção Educacional no Modelo de Regressão Quantílica

Os quantis de uma população ou amostra podem ser definidos da seguinte forma:

O quantil de ordem τ de uma população ou de uma amostra é o valor m tal que 100 $\tau\%$ dos valores populacionais ou amostrais são inferiores a ele, com $0 < \tau < 1$.

A definição para o caso populacional também pode ser feita utilizando a função de distribuição acumulada da variável aleatória X , em que:

$$F(x) = P(X \leq x)$$

Então utilizando a função inversa da distribuição acumulada no ponto τ , define-se que:

$$F^{-1}(\tau) = \inf\{x: F(x) \geq \tau\}$$

É o quantil de ordem τ da variável aleatória X . A mediana, nesse caso, seria definida como $F^{-1}(1/2)$. O primeiro e o terceiro quartil seriam $F^{-1}(1/4)$ e $F^{-1}(3/4)$, respectivamente.

Com essa definição para os quantis, podemos enunciar a ideia de regressão quantílica. O ponto inicial para a regressão quantílica é a função condicional quantílica (FCQ). Supondo que tenhamos uma variável aleatória continuamente distribuída, y_i , com função de densidade bem comportada. O FCQ no quantil τ dado um vetor, x_i , pode ser definida como:

$$Q_\tau = (Y_i|X_i) = F_Y^{-1}(\tau|X_i)$$

Em que $F_Y(y|X_i)$ é a função de distribuição para y_i condicional a X_i . Quando $\tau = .10$, por exemplo, $Q_\tau = (Y_i|X_i)$ descreve o menor decil de y_i dado X_i . Observando as mudanças no FCQ da nota dos alunos no Saeb em função das variáveis de background familiar e de infraestrutura das escolas, podemos mostrar se a dispersão na nota aumenta ou diminui com essas variáveis independentes.

O FCQ é a versão quantílica da função de expectativa condicional (FEC), derivada como a solução do problema do erro de predição quadrático médio:

$$E[Y_i|X_i] = \arg \min_{m(X_i)} E[(Y_i - m(X_i))^2]$$

Dessa forma, o FCQ, resolve o seguinte problema de minimização:

$$Q_i[Y_i|X_i] = \arg \min_{q(X_i)} E[\rho_\tau(Y_i - q(X_i))] \quad (3.2.1)$$

Em que $\rho_\tau(u) = (\tau - 1(u \leq 0))u$ é chamado de “check function” porque parece um sinal de correto quando colocado num gráfico. Quando $\tau = 0.5$, $Q_i[Y_i|X_i]$ é a mediana condicional, em que a mediana condicional minimiza os desvio absolutos.

De forma prática, a FCQ possui uma desvantagem em trabalhar com X_i contínuas ou de grande dimensão, sendo difícil fazer as estimações. No caso do estudo presente, trabalharemos com poucas variáveis independentes e em sua quase totalidade de

dummies. A regressão quantílica se finda ao substituímos um modelo linear por $q(X_i)$ em (3.2.1), produzindo:

$$\beta_\tau \equiv \arg \min_{q(X_i)} E[\rho_\tau(Y_i - X_i' b)] \quad (3.2.2)$$

O estimador da regressão quantílica, $\hat{\beta}_\tau$, é o análogo de (3.2.2) para distribuição amostral.

Ressalta-se que o vetor de parâmetros β deve ser indexado a τ , pois um dos interesses nesse estudo é exatamente estudar como esse vetor assume diferentes valores para percentis diferentes.

Diferentemente da análise de regressão em MQO, quando se faz a regressão pela média, a regressão quantílica possui uma característica peculiar, que é a quantidade de curvas a ser interpretada, mesmo considerando as mesmas variáveis explicativas ou não. Genericamente, quando os coeficientes estimados para β em função de τ variam em torno de uma constante muito proximamente, então se encontra evidência a favor da suposição de que os erros são independentes e identicamente distribuídos. Por outro lado, se esses coeficientes variam em função de τ , então os erros podem estar apresentando alguma forma de heterocedasticidade. Koeker (2005) mostra exemplos de ocasiões que podem ocorrer as duas situações de comportamento das curvas estimadas pela regressão quantílica. A conclusão é que os modelos de regressão quantílica possuem a capacidade de incorporar uma possível heterocedasticidade, que seria detectada a partir da variação das estimativas dos coeficientes para diferentes β a partir dos τ . A solução de $\hat{\beta}_\tau$ na sua versão de (3.2.2) pode ser encontrada usando métodos de programação linear.

De modo geral, o modelo de regressão quantílica utilizado aqui segue a estrutura do modelo MQO em (3.4), mantendo consigo a mesma hipótese de exogeneidade entre variáveis relacionadas à escola do aluno e de sua família.

4. Análise dos Resultados

Os resultados das estimações por MQO estão nas tabelas 4.1, 4.2 e 4.3 para as amostras dos alunos da base de dados do Saeb 2011. Essas regressões servirão de referência para as estimações por Regressão Quantílica que estão apresentadas nas tabelas 4.5, 4.6 e 4.7 e de forma completa no apêndice. O modelo MQO1 mostra a estimação do efeito médio do indicador de infraestrutura sem variáveis de controle. O modelo MQO2 estima o impacto do indicador de infraestrutura controlando para características do aluno e o *background* familiar. E o modelo MQO3 adiciona também *dummies* para região, tendo como referência a região Sudeste.

As tabelas 4.1, 4.2 e 4.3 indicam que o efeito da infraestrutura nas escolas tem efeito médio positivo e significativo sobre a proficiência dos alunos em matemática. Entretanto, a magnitude do impacto da infraestrutura reduz bastante com a adição dos controles, apesar da variável continuar significativa. Os impactos de um ponto no indicador de infraestrutura no modelo MQO3 são de 8,887, 6,080 e de 19,861 respectivamente para as classes de 5ª série, 9ª série e terceiro ano do ensino médio, respectivamente. Esses resultados são equivalentes a 18,68%, 13,11% e 31,25% dos desvios padrão das proficiências médias em matemática para os alunos de 5ª série, 9ª série e 3º ano, respectivamente. Esse resultado é compatível com o trabalho de Gleww et al. (2013), que estuda, entre outras coisas, a importância da infraestrutura usando a meta-análise, e Hanushek et al. (2008), que conclui que o impacto da infraestrutura na queda da taxa de evasão escolar é positivo.

Para os estudantes da 5ª série, o efeito da infraestrutura tem magnitude média comparada às outras variáveis controladas no modelo MQO3. Tanto que, se fizermos uma comparação entre um aluno que estuda no Sudeste e outro que estuda no Nordeste e dermos ao aluno do Nordeste uma infraestrutura escolar adequada e ao aluno do Sudeste uma infraestrutura deficitária, mesmo assim o aluno do Sudeste conseguiria tirar uma nota 10 pontos acima na média. Isso evidencia a importância também da infraestrutura que está fora da escola, como saneamento básico, sistema de saúde e a cultura local. Para as séries seguintes, essa mesma relação diminui, chegando a praticamente se compensarem no terceiro ano do ensino médio. Por outro lado, a importância da infraestrutura escolar aumenta para a 9ª e para o 3º ano em comparação com as outras variáveis controladas.

É interessante revisitar também algumas variáveis de controle usualmente utilizadas na literatura. Com relação ao *background* do aluno como estudante, o fato de ele já haver repetido uma ou mais vezes alguma série anterior resulta numa nota média menor e afeta especialmente mais aos alunos de 3º ano, como se as deficiências dos

alunos repetentes não fossem não somente sanadas de maneira adequada, mas também acumuladas ao longo do tempo.

Um resultado interessante dessa pesquisa é que o impacto médio “Creche” é menor do que o impacto da variável “Pré-escola” para todas as séries estudadas, porém ambos são positivos e significantes (na 9ª série o efeito é positivo e significativo apenas após a inclusão dos controles das regiões para a variável “Creche”). Esse resultado é corroborado pelo estudo feito por Heckman et al. (2010b), que estuda a importância do programa “The Perry Preschool Program” no aumento de salários, no rendimento escolar, na redução da criminalidade e na redução de custos para o estado em programas de auxílio do governo. O programa era aplicado para crianças de 3 a 4 anos de idade. Entretanto, contrapõem de certa forma inúmeros artigos que analisam a importância da creche para o desenvolvimento cognitivo, como Duncan (2003).

A educação dos pais representa bem o *background* familiar e possui impacto positivo, como se esperava, no rendimento dos alunos. Curiosamente, quanto maior a série, maior é a magnitude da variável, e ela aumenta para todas as séries depois de fazer o controle por região. Outra variável de *background* familiar importante é a quantidade de pessoas que moram na mesma casa do aluno. O resultado já seria o esperado, já que com famílias menores os recursos familiares podem ser direcionados de maneira maior à criança.

Duas variáveis no modelo que são alvos de políticas públicas são “Computador com internet em casa” (ex: “Computador para Todos” (lei 11.196/05)) e “Trabalha fora de casa” (ex: decreto 6481, de 12.06.08). Esta variável possui efeito negativo e significativo para alunos de todas as séries, sendo que sua magnitude é maior para alunos de 5ª série e do 3º ano. Já para aquela, o impacto médio é positivo e significativo para todas as séries. É claramente possível que ambas as variáveis sofram em alguma medida de algum viés, por estas estarem muito provavelmente ligadas a fatores de renda da família. Mesmo assim, os sinais dos coeficientes dessas variáveis não trazem surpresa. A variável que representa o trabalho doméstico (Faz trabalho doméstico 2 horas ou mais por dia), embora não seja alvo de políticas públicas, apresenta resultados semelhantes ao trabalho fora de casa e evidencia o impacto da falta de recursos no domicílio envolto ao estudante.

Por fim, no que se refere às variáveis de controle, três variáveis que indicam o potencial do esforço do aluno sobre a resultado da proficiência em matemática é se ele faz sempre o dever de casa de matemática, se ele sempre lê livros, e se ele frequenta sempre bibliotecas. A primeira variável mostra efeito positivo, alto e significativo, para todas as séries analisadas, mudando sua magnitude muito pouco mesmo após o controle das regiões. A segunda variável possui impacto positivo, pequeno e significativo para as 5ª e 9ª séries e negativo, pequeno e significativo para o

3º ano do ensino médio. Já a terceira variável, que só é disponível a ser respondida para alunos da 5ª e 9ª série, apresenta sinais negativos e significantes dos coeficientes.

Esse último resultado não foi o que era esperado e mostra como é forte o problema da endogeneidade. É muito provável que haja variáveis omitidas no modelo que estão enviesando os resultados. Nas tabelas 3.2 e 3.3, o hábito de frequentar a biblioteca é maior para os percentis mais baixos da distribuição dos alunos pela nota, o que teoricamente é difícil de explicar sem uma análise mais profunda.

Tabela 4.1-Impacto Médio da Proficiência em Matemática-5ª Série

	MQO1	MQO2	MQO3
Infraestrutura	24,878*** (0,377)	16,526*** (0,360)	8,887*** (0,383)
Branco		6,120*** (0,319)	4,186*** (0,318)
Tem computador com internet em casa		11,503*** (0,309)	8,091*** (0,312)
Mora com 4 ou mais pessoas		-12,415*** (0,304)	-11,416*** (0,300)
Lê livros		2,162*** (0,301)	2,460*** (0,296)
Frequenta bibliotecas		-6,615*** (0,364)	-6,987*** (0,341)
Faz Trabalhos Domésticos 2 horas ou mais por dia		-6,7*** (0,290)	-6,297*** (0,285)
Trabalha fora de casa		-12,563*** (424)	-11,537*** (0,418)
Já foi reprovado		-21,021*** (0,323)	-19,326*** (0,320)
Faz dever de matemática		13,483*** (0,341)	13,216*** (0,336)
Creche		4,796*** (0,372)	5,528*** (0,368)
Pré-escola		10,665*** (0,372)	10,525*** (0,367)
Pai ou mãe tem ensino superior		1,418*** (0,409)	1,954*** (0,403)
Norte			-14,408*** (0,514)
Nordeste			-18,972*** (0,393)
Sul			1,676*** (0,430)
Centro Oeste			-4,465*** (0,555)
Constante	205,520*** (0,156)	216,056*** (0,516)	213,275*** (0,545)
R-quadrado	0,045	0,212	0,236

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Tabela 4.2-Impacto Médio da Proficiência em Matemática-9ª Série

	MQO 1	MQO 2	MQO 3
Infraestrutura	16,870*** (0,475)	9,832*** (0,586)	6,080*** (0,548)
Branco		8,527*** (0,396)	6,260*** (0,393)
Tem computador com internet em casa		9,585*** (0,381)	7,300*** (0,390)
Mora com 4 ou mais pessoas		-9,019*** (0,363)	-8,34*** (0,362)
Lê livros		0,921** (0,416)	1,530*** (0,414)
Frequenta bibliotecas		-5,041*** (0,560)	-4,73*** (0,557)
Faz Trabalhos Domésticos 2 horas ou mais por dia		-1,135*** (0,645)	-1,15*** (0,361)
Trabalha fora de casa		-0,768* (0,465)	-1,26*** (0,462)
Já foi reprovado		-19,44*** (0,396)	-18,63*** (0,397)
Faz dever de matemática		14,004*** (0,369)	14,160*** (0,370)
Creche		0,479 (0,385)	2,130*** (0,511)
Pré-escola		7,229*** (0,477)	7,900*** (0,477)
Pai ou mãe tem ensino superior		6,894*** (0,528)	7,170*** (0,584)
Norte			-8,17*** (0,718)
Nordeste			-9,86*** (0,494)
Sul			6,280*** (0,516)
Centro Oeste			-0,51 (0,713)
Constante	243,665*** (0,177)	239,094*** (0,618)	241,640*** (0,658)
R-quadrado		0,016	0,146
			0,158

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Tabela 4.3-Impacto Médio da Proficiência em Matemática-3º Ano

	MQO 1	MQO 2	MQO 3
Infraestrutura	38,332*** (0,323)	20,584*** (0,435)	19,861*** (0,433)
Branco		12,433*** (0,586)	8,224*** (0,610)
Tem computador com internet em casa		20,646*** (0,652)	17,466*** (0,660)
Mora com 4 ou mais pessoas		-5,346*** (0,562)	-3,948*** (0,559)
Lê livros		-3,128*** (0,557)	-1,974*** (0,550)
Faz Trabalhos Domésticos 2 horas ou mais por dia		-14,217*** (0,569)	-14,011*** (0,563)
Trabalha fora de casa		-6,361*** (0,648)	-8,313*** (0,647)
Já foi reprovado		-29,618*** (0,654)	-28,615*** (0,650)
Faz dever de matemática		11,693*** (0,557)	11,632*** (0,552)
Creche		6,580*** (0,847)	8,630*** (0,854)
Pré-escola		10,594*** (0,778)	11,056*** (0,774)
Pai ou mãe tem ensino superior		22,155*** (0,686)	23,114*** (0,683)
Norte			-18,948*** (0,874)
Nordeste			-15,771*** (0,835)
Sul			1,535* (0,818)
Centro Oeste			-8,861*** (0,852)
Constante	278,930*** (0,249)	268,483*** (1,025)	279,539*** (1,139)
R-quadrado		0,171	0,367
			0,38

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

A tabela 4.4 mostra os resultados do teste Breusch-Pagan. Note que o p-valor é 0,00, o que infere que a modelagem em MQO contém heterocedasticidade. Desta forma, o caminho é agora testar outro método que trate do problema da variância não constante ao longo da distribuição.

Tabela 4.4-Teste Breusch-Pagan

	Chi ²	(K)	P-valor
Quinta Série	1200,93	17	0,00
Nona Série	312,02	17	0,00
Terceiro ano	234,34	16	0,00

A utilização nos modelos MQO3 das tabelas 4.1, 4.2 e 4.3 dos erros padrões robustos é justificada pela presença de heterocedasticidade. Esse mecanismo permite o uso de estatísticas para o modelo, como o teste t. Parte-se então para a regressão quantílica com base no modelo MQO3. Esse método torna possível avaliar o impacto da infraestrutura no desempenho dos alunos ao longo da distribuição para a 5ª série, 9ª série e 3º ano do ensino médio.

O resultado das tabelas 4.5, 4.6 e 4.7 mostram que o efeito de um ambiente escolar com melhor infraestrutura impacta de modo diferente os percentis de distribuição, sendo que essa mudança no coeficiente não muda da mesma maneira entre os alunos de diferentes séries. Para os alunos de 5ª série da faixa de distribuição mais baixa, um ponto a mais no indicador de infraestrutura significa aproximadamente 5,1 pontos a mais na nota em matemática. Esse impacto vai aumentando continuamente ao longo da distribuição de forma que no maior quantil, o efeito é mais que o dobro (12,32 pontos) em comparação ao primeiro quantil.

O fato do impacto da infraestrutura escolar aumentar ao longo da distribuição é interessante porque evidencia a importância desse insumo para o resultado dos alunos e de certo modo não substitui o *background* familiar, visto que nas tabelas 3.2, 3.3 e 3.4 vê-se que os alunos com as melhores notas na média possuem *background* familiar maior. Se fosse o inverso, a resposta seria que a infraestrutura compensa o déficit em dotações que os estudantes enfrentam, enquanto que no final das contas a infraestrutura pouco importaria para os alunos com melhores notas e com melhores condições familiares na média.

Observando melhor a hipótese de que a infraestrutura escolar tem impacto maior sobre os estudantes de quinta série de maior *background* familiar, foi feita a seguinte análise. Foi selecionados os alunos que tiveram pouco *background* familiar. Para isso, admitimos que o fato do estudante tenha entrado já na pré-escola supõe uma preocupação dos pais com a educação dos filhos. Nesse caso, para aqueles alunos de quinta série que não tiveram acesso à pré-escola e que tiraram uma proficiência baixa

em matemática (menor do que o terceiro decil, tabela 3.1), a regressão simples entre a proficiência dos alunos e o indicador de infraestrutura gerou um impacto na nota de 2,33 pontos. Depois, foram selecionados os alunos de quinta série com elevado background familiar (alunos que tenham frequentado a pré-escola e que um dos pais tenha ensino superior completo) e que tenham tirado uma proficiência alta em matemática (acima do sétimo decil, tabel 3.1). A mesma regressão indicou que um ponto a mais de infraestrutura geraria um aumento de 7,65 pontos.

Para os alunos da nona série o efeito de um ponto a mais no indicador de infraestrutura é muito parecido ao longo da distribuição e variando sempre ao redor do coeficiente de 6 pontos. O que é interessante desse nível de ensino, é que embora não haja grandes diferenças do efeito da do indicador de infraestrutura na nota dos estudantes, a magnitude do impacto é consideravelmente menor se comparado com as outras séries analisadas. Nesse caso, apenas a prática de fazer o dever de casa gera um resultado muito maior em todos os percentis de nota.

Já para os estudantes do 3º ano o efeito da infraestrutura cresce cerca de 1,3 ponto até o terceiro decil (de 19,57 pontos no primeiro quantil para 20,84 no terceiro) e depois decresce continuamente até reduzir o efeito em pouco menos de 3 pontos (chegando a 17, 97 pontos no último quantil). Embora de fato não haja uma grande diferença entre os coeficientes mostrados na Figura 3 o aspecto relevante é que estudantes com alto desempenho na proficiência em matemática dependem menos da infraestrutura escolar. Alunos de terceiro ano geralmente estão em fase de pré-vestibular, isso implica que o esforço do aluno é preponderante para o seu sucesso no vestibular. Tanto é que a experiência de já ter reprovado, o que poderia ser uma variável proxy para o seu comprometimento como aluno, diminui a nota em matemática do alunos numa intensidade muito maior do que nas outras séries (isso também poderia ser interpretado como um déficit acumulado ao longo dos anos e a não recuperação do estudante após passar duas vezes pela mesma série). O fato é que os alunos, principalmente aqueles que estão disputando as vagas nas melhores universidades intensificam o esforço próprio. Isso explica o porquê da infraestrutura ser menos importante entre os estudantes de ponta do os outros que estão na média.

As estimativas completas das tabelas 4.5, 4.6 e 4.7 se encontram nas tabelas A1 até A9. Vale destacar algumas variáveis de controle. Como nas regressões em MQO, em geral, os sinais e a significância das variáveis se mantiveram nas regressões quantílicas, salvo a variável “Trabalha fora de casa” para alunos da nona série, em que seu coeficiente começa negativo e significativo e a partir da mediana ele se torna negativo não significativo e depois positivo não significativo.

Segundo as estimações, fazer dever de casa é bom para os alunos em todos os quantis de distribuição, de todas as séries, mas é melhor para os alunos dos quantis superiores. Quanto maior é o quantil de distribuição ao qual o aluno pertence, mais ele

se destaca com relação a seus pares ao fazer dever de casa de matemática sempre. Esse fato ocorre para as três séries investigadas. Mas se destaca na quinta série, em que o diferencial de rendimento dobra entre os quantis extremos (8,18 para o primeiro quantil e 16,33 para o último quantil).

Outras variáveis que tem seu coeficiente positivo aumentado quase que continuamente nos quantis de distribuição maiores são “Começou na Creche” e “Começou na Pré-escola”, sendo que só para os alunos de quinta série em que o coeficiente é de 4,05 e 7,94 para os alunos do primeiro quantil (6,3 e 11,1 para o nono quantil) que começaram na creche e os que começaram na pré-escola, respectivamente. Para o terceiro ano apenas a variável “Começou na Creche” cresce ao longo da distribuição (saindo de 4,78 no primeiro quantil para 9,64 no nono quantil). No nono ano, os coeficientes se mantêm próximos.

Algo importante a notar, que também foi verificado nas regressões em MQO, é a importância da educação dos pais para os estudantes das diferentes séries. Quanto mais elevada é a série, maior a importância da educação dos pais. Na quinta-série, tanto na regressão em MQO quanto na regressão quantílica para os diversos quantis de distribuição, o fato de o estudante ter pai ou mãe com um diploma universitário aumenta a nota em no máximo 2,6 pontos (último quantil da regressão quantílica). Na nona série, o coeficiente da regressão em MQO completa é de 7,17, enquanto que na regressão quantílica o beta sai de 2,85 no primeiro decil aumentando continuamente até 10,32 no último decil. Já no terceiro ano o valor do coeficiente é acima de 20 pontos tanto para a regressão em MQO, no modelo MQO2 e MQO3, quanto para a regressão quantílica, em toda a sua distribuição.

Nas tabelas 2, 3 e 4, onde são mostradas características descritivas das variáveis do modelo, vê-se que, ao contrário do que ocorre para estudantes da oitava série e terceiro ano quando a proporção de alunos com pais com diploma de ensino superior aumenta quando cresce o percentil de distribuição, essa proporção não se altera significativamente entre os diversos quantis de distribuição para os estudantes de quinta série.

A teoria da função de produção educacional tem como principal referência como indicador para o *background* familiar exatamente a educação dos pais. Não que o *background* familiar não seja importante e não demonstre que há diferenças entre as características médias observadas dos estudantes que estão no estrato acima em comparação com os que estão no estrato abaixo, como bem demonstra as outras variáveis que também indicam o suporte familiar, como “Começou na Pré-escola”, “Trabalha fora de Casa”, “Mora com 4 ou mais pessoas”. Mas, pelo menos nesse estudo, para essa amostra (quinta série) o grau de educação dos pais não demonstrou tanta importância no desempenho dos estudantes.

As figuras 1, 2 e 3 mostram o comportamento do impacto da infraestrutura ao longo da distribuição com o intervalo de confiança. Notando que a curva na quinta série é crescente, o que pode demonstrar que a infraestrutura provoca uma ampliação da desigualdade entre os alunos com as melhores notas. Na nona série ela não demonstra um movimento muito claro, oscilando perto de 6 pontos e no terceiro ano a curva cresce, mas depois da mediana cai novamente. Este último caso pode ter haver com a questão da maturidade dos alunos de terceiro ano. Para os alunos melhores, a infraestrutura passa a ter um peso cada vez menor no desempenho no teste.

Tabela 4.5-Impacto da Infraestrutura da escola nas Notas de Acordo com os Quantis de Distribuição-5ª série

	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Intercepto	160,67*** (1,00)	177,56*** (0,63)	190,24*** (0,62)	201,45*** (0,68)	212,49*** (0,59)	223,34*** (0,65)	234,13*** (0,68)	248,46*** (0,84)	268,47*** (0,94)
Infraestrutura	5,12*** (0,45)	7,13*** (0,48)	7,90*** (0,52)	8,18*** (0,52)	8,43*** (0,53)	9,73*** (0,53)	10,38*** (0,45)	11,53*** (0,68)	12,31*** (0,85)
Controle alunos, Famílias e Região	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Tabela 4.6-Impacto da Infraestrutura da escola nas Notas de Acordo com os Quantis de Distribuição-9ª série

	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Intercepto	187,75*** (0,90)	205,6*** (0,93)	219,00*** (0,85)	230,37*** (0,89)	241,67*** (0,90)	252,77*** (0,71)	264,72*** (0,82)	278,15*** (0,87)	295,82*** (1,09)
Infraestrutura	5,87*** (1,17)	6,67*** (0,82)	5,90*** (0,79)	6,55*** (0,71)	6,70*** (0,66)	6,53*** (0,62)	6,80*** (0,58)	6,65*** (0,58)	5,73*** (0,94)
Controle alunos, Famílias e Região	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Tabela 4.7-Impacto da Infraestrutura da escola nas Notas de Acordo com os Quantis de Distribuição-3º ano

	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Intercepto	204,62*** (4,03)	231,00*** (3,15)	247,55*** (2,62)	257,68*** (2,97)	271,21*** (2,42)	283,39*** (2,05)	297,72*** (2,41)	311,24*** (2,60)	326,41*** (2,29)
Infraestrutura	19,57*** (0,85)	20,25*** (0,59)	20,84*** (0,60)	21,11*** (0,53)	21,03*** (0,49)	20,57*** (0,45)	20,33*** (0,48)	19,34*** (0,55)	17,97*** (0,60)
Controle alunos, Famílias e Região	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Figura 1- Impacto da Infraestrutura Escolar na Proficiência escolar ao Longo da Distribuição de Conhecimento do Aluno de 5ª Série.

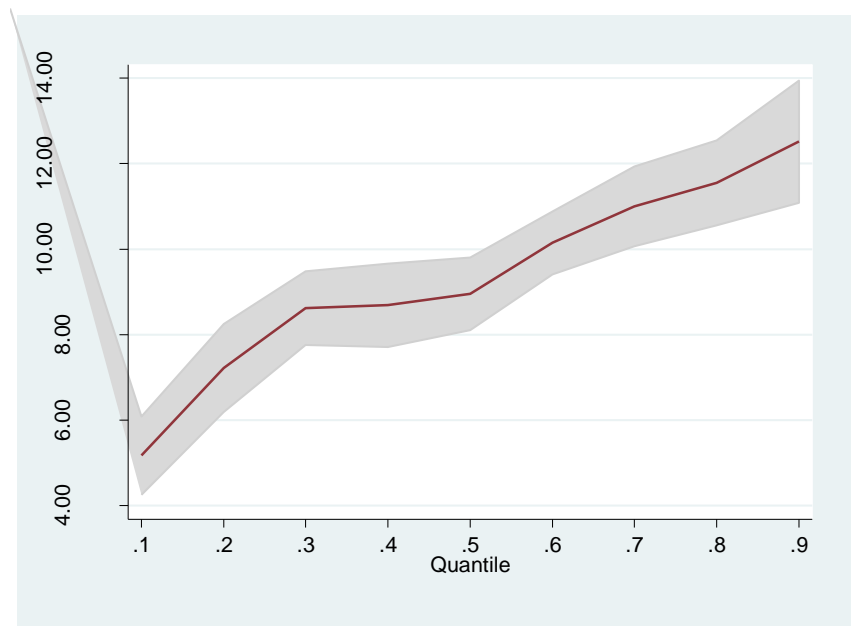


Figura 2- Impacto da Infraestrutura Escolar na Proficiência escolar ao Longo da Distribuição de Conhecimento do Aluno de 9ª Série.

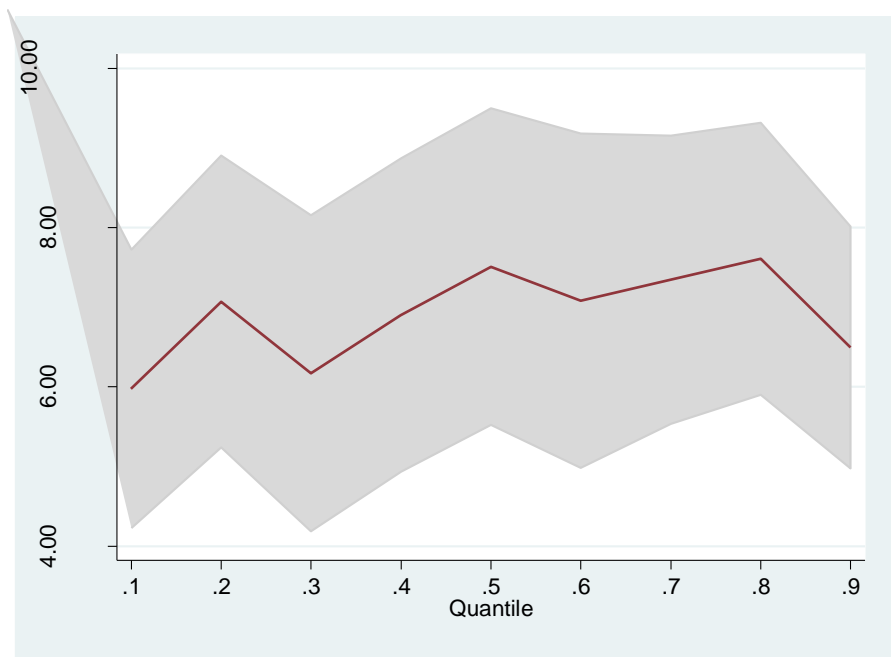
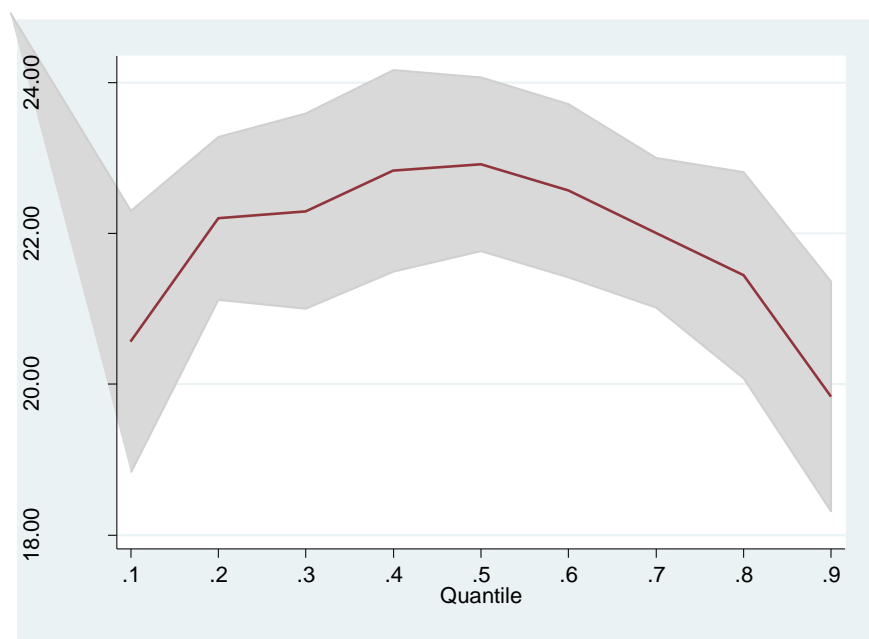


Figura 3- Impacto da Infraestrutura Escolar na Proficiência escolar ao Longo da Distribuição de Conhecimento do Aluno de 3º Ano.



5. Conclusão

O objetivo deste trabalho é estimar o impacto da infraestrutura nas escolas de ensino primário e secundário de escolas públicas do Brasil. A partir de seus resultados, pretende-se ser mais uma evidência empírica no auxílio de políticas públicas voltadas para a educação. A literatura, através da meta-análise, indica que tanto a área de infraestrutura física, quanto à área de infraestrutura organizacional tende a dar uma importância positiva no rendimento dos estudantes, embora não seja consensual.

Para atingir esse objetivo, foi montado um indicador de infraestrutura composto de diversas variáveis relacionadas à infraestrutura escolar. Após a montagem deste indicador, foi estimado o impacto do indicador infraestrutura através de MQO para as 5ª e 9ª séries do ensino fundamental e para o 3º ano do ensino médio, controlando o por variáveis que representam o *background* familiar, as características dos alunos, e as regiões do país.

No entanto, um das hipóteses iniciais era de que esse efeito era heterógeno ao longo da distribuição da nota e que o efeito do indicador seria maior para os quantis menores para todas as séries. Portanto, também foi estimado o modelo de Regressão Quantílica para a mesma especificação utilizada no modelo em MQO.

Os resultados mostram que o indicador de infraestrutura tem um impacto médio positivo e significativo para proficiência em matemática e para todas as três séries analisadas. O mesmo ocorre para todas as séries quando analisamos os quantis. No caso da 5ª série, o impacto do indicador de infraestrutura cresce continuamente ao longo da distribuição, intuindo que a construção de infraestrutura nas escolas tende a aumentar o diferencial de nota entre os alunos do início e do topo da distribuição da nota. Para os alunos da 9ª série, Não houve mudanças significativas do impacto do indicador ao longo da distribuição. Já no 3º ano, o impacto foi maior para os quantis centrais da distribuição, sendo menor o impacto no 9º quantil do que no primeiro. Esse resultado indica que os melhores alunos trazem uma bagagem suficiente para que a infraestrutura escolar não seja algo tão preponderante.

Estudos importantes ligados à educação inferem que é de suma importância investir principalmente na educação em níveis inferiores, como a pré-escola (ver, por exemplo, Heckman (2010)). Crianças que tiveram um maior suporte educacional dos pais e da escola quando menores podem naturalmente ser mais receptivos à influência de novas variáveis, como infraestrutura, o que faz parte da aptidão que eles adquirem nos anos iniciais de vida. Isso explica porque os alunos de quinta-série nas distribuições de cima da nota são mais influenciados pela infraestrutura escolar, já que os alunos dos percentis superiores também são aqueles que também têm maior “background” familiar.

O inverso ocorre para os estudantes de terceiro ano, ou seja, a infraestrutura é menos importante para aqueles que estão nos percentis superiores. Não é que a lógica explicada para o caso da quinta-série muda para os alunos do terceiro ano, é que uma nova variável entra nessa explicação, o pré-vestibular. Essa variável pode fazer com que o esforço próprio do estudante minimize, pelo menos entre os estudantes de ponta, a importância da infraestrutura.

Mesmo que os resultados encontrados sejam potencialmente relevantes, melhorias no modelo devem ser implementadas com o objetivo de lidar com o problema de endogeneidade, virtualmente causada, nesse caso, por simultaneidade entre as variáveis explicativas e a variável explicada e por omissão de variáveis explicativas.

6. Referências Bibliográficas

ANDERSON, M. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". **Berkeley University**. July, 2007.

ARROW, K. "Models of Job Discrimination" and "Some Mathematical Models of Race in the Job Market", in M.H. Pascal, ed. *Racial Discrimination in Economic Life*, **Lexington**, 1972.

CARD, DAVID AND ALAN B. KRUEGER. "Does School Quality Matter? Returns to Education and School Characteristics of Public Schools in the United States", **Journal of Political Economy** 100 (1992a), 1-40.

COLEMAN, J.S., CAMPBELL, E.Q., HOBSON, C.J., MCPARTLAND, J., MOOD, A.M., WEINFELD, F.D. AND YORK, R.L. (1966). *Equality of Educational Opportunity*. **Washington, D.C.: U.S. Government Printing Office**, 1966.

BETTS, J. AND SHKOLNIK. "The Behavioral Effects of Variations in Class Size: The Case of Math Teachers". **Educational Evaluation and Policies Analysis** June 20, 1999 vol. 21 no. 2 193-213.

DATAR, A. AND MASON, B. Do reductions in class size "crowd out" parental investment in education? . **American of Education Review**. 27 (2008) 712-723.

GLEWWE, PAUL AND MICHAEL KREMER. 2006. "Schools, Teachers, and Educational Outcomes in Developing Countries." In **Handbook of the Economics of Education**, ed. ERIC A. HANUSHEK and FINIS WELCH, 943-1017. Amsterdam: North Holland.

GLEWWE, PAUL, MICHAEL KREMER, SYLVIE MOULIN AND ERIC ZITZEWITZ. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." **Journal of Development Economics** 74 (1): 251-68.

GLEWWE, ET AL. (2013). "Education Policy in Developing Countries". **The University of Chicago Press**.

LINDEN, LEIGH. 2008. "Complement or Substitute? The Effect of Technology on Student Achievement in India." **JPAL Working Paper, MIT**.

HANUSHEK ET AL. "Aggregation and Estimation Effect of School Resources". **The Review of Economics Statistics**, Vol. 78, Issue 4 (1996), 611-627.

- HANUSHEK, ERIC A. 1979. "Conceptual and empirical issues in the estimation of educational production functions." **Journal of Human Resources** 14,no.3 (Summer):351-388.
- HANUSHEK, E.A. and KINKO D.D. (2000). 'Schooling, labor force quality, and the growth of nations', **American Economic Review**, vol. 90 (5) (December), pp. 1184–208.
- HANUSHEK, E.A. (1986). 'The economics of schooling: production and efficiency in public schools', **Journal of Economic Literature**, vol. 24 (3) (September), pp. 1141–77.
- HANUSHEK, E.A. (1986). 'The economics of schooling: production and efficiency in public schools', **Journal of Economic Literature**, vol. 24 (3) (September), pp. 1141–77.
- HANUSHEK, E.A. (1996). 'A more complete picture of school resource policies', **Review of Educational Research**, vol. 66 (3) (Fall), pp. 397–409.
- HANUSHEK, E.A. (1979). 'Conceptual and empirical issues in the estimation of educational production functions', **Journal of Human Resources**, vol. 14 (3) (Summer), pp. 351–88.
- HANUSHEK, ERIK. The Trade-off Between Child Quantity and Quality. **Journal of Political Economy**,100(1), February 1992, pp. 84-117.
- HANUSHEK, Erik. Alternative School Policies and the Benefits of General Cognitive Skills. *Economics of Education Review*, 25(4), August 2006, pp. 447-462.
- HANUSHEK, E. "The Failure of Input-Based Schooling Policies". **The Economic Journal**, 113 (February), F64–F98. _ Royal Economic Society 2003.
- HANUSHEK, ERIC A., VICTOR LAVY AND KOHTARO HITOMI. 2008. "Do Students Care about School Quality? Determinants of Dropout Behavior in Developing Countries." **Journal of Human Capital** 1 (2): 69-105.
- HANUSHEK, ERIC A. AND LUDGER WOESSMANN. 2008. "The Role of Cognitive Skills in Economic Development." **Journal of Economic Literature** 46 {3}: 607-68.
- HECKMAN, J. J., S. H. MOON, R. PINTO, P. A. SAVELYEV, and A. Q. YAVITZ (2010b, February). The rate of return to the HighScope Perry Preschool Program. **Journal of Public Economics** 94 (1-2), 114{128.
- HENRY LEVI. "A New Model of School Effectiveness." *Do teachers make differences?* **U.S Office of Education**, 1971.
- JAMES J. HECKMAN and DIMITRIY V. MASTEROV. "The Productivity Argument for Investing in Young Children". **Review of agricultural Economics**, Vol. 29, Number 3, 446-493.

- KATZMAN, M. "The Political Economy of Urban Schools". **Cambridge**, 1971
- KRUEGER, A.B. (1999). 'Experimental estimates of education production functions', **Quarterly Journal of Economics**, vol. 114 (2) (May), pp. 497–532.
- MURNANE, R.J., WILLETT, J.B., BATAATZ, M.J. AND DUHALDEBORDE, Y. (2001). "Do different dimensions of male high school students skills predict labor market success a decade later? Evidence from the NLSY", **Economics of Education Review**, vol. 20 (4) (August), pp. 311–20.
- NEAL, D.A. AND JOHNSON, W.R. (1996). "The role of pre-market factors in black-white differences", **Journal of Political Economy**, vol. 104 (5) (October), pp. 869–95.
- OECD, 2011. **PISA 2012 Results: What Makes Schools Successful (Volume IV)**. http://www.oecd-ilibrary.org/education/pisa_19963777.
- SCHWEINHART, L. J., J. MONTIE, Z. XIANG, W. S. BARNETT, C. R. BEL_ELD, AND M. NORES (2005). *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti, MI: **High/Scope Press**.
- STIGLITZ, J. The Theory of "Screening", Education, and the Distribution of Income. **The American Economic Review**. Vol. 65, n°3 (1975) 283-300.
- SPENCE, M. "Job Market Signalling", **Quart. J. Econ.**, 1973, 355-379.
- SUMMERS AA, WOLFE BL. 1977. Do schools make a difference? **American Economic Review** 67(4): 639–652.
- WOLPIN, K AND TODD, P (2003). "On the Specification and Estimation of the Production Function for Cognitive Achievement". **The Economic Journal**, 113 (February), F3–F33. _ **Royal Economic Society** 2003.

Apêndice

Apêndice 1-Resultado da Regressão Entre Variáveis de Infraestrutura Contra Nota da Escola

Mínimos Quadrados Ordinários	Quinta Série		Nona Série		Terceiro Ano	
	Coefficiente	Erro padrão	Coefficiente	Erro padrão	Coefficiente	Erro padrão
Sala de aula em bom estado	0,034***	,010	-0,031**	,013	0,025	,026
Banheiros em bom estado	0,022**	,010	-0,026*	,014	0,11***	,029
Instalação elétrica em bom estado	-0,013	,010	-0,122***	,013	0,08***	,028
Sinais de depredação	-0,04***	,009	0,009	,011	-0,05***	,021
Computadores para os alunos	0,105***	,009	0,069***	,012	-0,003	,024
Computadores para uso administrativo	0,037***	,010	0,038***	,013	0,05**	,029
Máquina Copiadora	0,112***	,009	0,072***	,012	0,107***	,025
Biblioteca	0,085***	,010	0,235***	,014	-0,035**	,027
Quadra de esportes	0,018**	,010	0,038***	,013	0,044***	,024
laboratório	-0,207***	,009	-0,042***	,012	0,049***	,021
Sala de música	-0,123***	,010	-0,145***	,013	0,067***	,016
Sala de artes	0,001	,010	-0,023**	,013	0,113	,016
Sala de leitura	-0,082***	,010	-0,199***	,013	0,021	,023
Programa de redução de taxa de reprovação	0,071***	,009	0,06***	,012	-0,029***	,019
Programa de reforço escolar	-0,134***	,009	0,055***	,012	-0,119***	,020
Não ocorreu problemas financeiros	-0,047***	,009	-0,135***	,011	0,046***	,021
Não ocorreu falta de professor	0,111***	,009	-0,158***	,012	0,126***	,022
Acessibilidade	-0,028***	,009	-0,08***	,011	0,008	,022
Mesas para cadeirantes	-0,024***	,008	-0,031***	,011	0,018	,017
Atividades esportivas	-0,106***	,009	0,088***	,012	0,007***	,030
Atividades artísticas	-0,016***	,009	-0,03***	,012	0,072	,026
R-quadrado	0,074		0,15		0,043	

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 2- Regressão Quantílica (1º decil)

	Quinta Série		Nona Série		Terceiro Ano	
	Coefficiente	Erro Padrão	Coefficiente	Erro Padrão	Coefficiente	Erro Padrão
Infraestrutura	5,125617***	0,45	5,878194***	1,17	19,57459***	0,86
Trabalha fora de casa	-11,17207***	0,41	-4,916021***	0,92	-9,005578***	1,03
Frequenta biblioteca	-5,37559***	0,54	-6,494855***	1,08	-	-
Faz dever de casa de matemática	8,155873***	0,45	11,73746***	0,61	9,858181***	0,97
Lê livros	2,684676***	0,41	1,162246	0,77	-1,166878	0,84
Pais ou mãe tem ensino superior	2,130221***	0,62	2,851279***	1,22	21,47723***	1,61
Começou na creche	4,056835***	0,50	1,270617	0,81	4,78653***	1,54
Começou na pré-escola	7,990752***	0,52	8,385401***	0,69	10,9079***	1,04
Aluno é Branco	2,401947***	0,42	6,746102***	0,83	10,10718***	1,05
Já foi reprovado	-12,47105***	0,43	-17,47961***	0,64	-25,37221***	1,15
Mora com 4 ou mais pessoas em casa	-8,407036***	0,41	-7,919287***	0,64	-4,75286***	0,95
Faz trabalho doméstico	-6,113855***	0,38	-1,283788**	0,64	-9,124525***	1,06
Tem computador com internet em casa	5,932642***	0,45	5,780071***	0,58	10,67302***	1,22
Norte	-8,738386***	0,77	-5,90136***	1,05	3,702726**	3,70
Nordeste	-12,70953***	0,62	-7,122698***	0,86	3,886869	3,89
Sul	4,225022***	0,58	10,65216***	0,85	13,76696***	2,75
CentroOeste	-2,004668***	0,83	1,972779***	1,08	3,382839	3,38
Constante	160,6771***	1,00	187,7555***	0,90	204,6291***	4,04

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 3- Regressão Quantílica (2º decil)

	Quinta Série		Nona Série		Terceiro Ano	
	Coefficiente	Erro Padrão	Coefficiente	Erro Padrão	Coefficiente	Erro Padrão
Infraestrutura	7,13***	0,48	6,671288***	0,82	20,25744***	0,60
Trabalha fora de casa	-11,62083***	0,45	-3,510518***	0,80	-9,079594***	0,85
Frequenta biblioteca	-6,474431***	0,39	-6,452048***	0,86	-	-
Faz dever de casa de matemática	9,924628***	0,38	12,87406***	0,54	9,966941***	0,86
Lê livros	2,829466***	0,39	1,289506***	0,64	-1,671231*	0,88
Pais ou mãe tem ensino superior	1,678878***	0,49	4,732549***	0,89	23,39051***	1,14
Começou na creche	4,390475***	0,50	2,190835***	0,72	8,364158***	1,15
Começou na pré-escola	8,906669***	0,49	8,461367***	0,66	12,28898***	1,10
Aluno é Branco	3,107108***	0,40	6,152138***	0,62	9,856331***	0,89
Já foi reprovado	-14,71444***	0,39	-18,1251***	0,57	-29,53059***	1,09
Mora com 4 ou mais pessoas em casa	-10,12554***	0,31	-8,431486***	0,60	-3,588089***	0,73
Faz trabalho doméstico	-6,592891***	0,31	-0,603565	0,47	-11,93539***	0,89
Tem computador com internet em casa	6,847607***	0,42	7,003623***	0,48	14,17258***	0,97
Norte	-10,7912***	0,54	-7,342614***	0,99	-11,9549***	2,78
Nordeste	-14,75155***	0,48	-9,000423***	0,66	-9,602967***	2,83
Sul	3,54918***	0,48	9,180162***	0,69	11,1728***	2,25
CentroOeste	-2,735512***	0,59	-0,322003	1,02	-2,238504	2,53
Constante	177,5608***	0,63	205,5802***	0,93	231,0345***	3,16

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 4- Regressão Quantílica (3º decil)

q30	Quinta Série		Nona Série		Terceiro Ano	
	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão
Infraestrutura	7,904741***	0,52	5,9094***	0,79	20,84942***	0,61
Trabalha fora de casa	-11,32963***	0,53	-1,864285***	0,69	-8,796804***	0,79
Frequenta biblioteca	-6,719315***	0,40	-6,103234***	0,80	-	-
Faz dever de casa de matemática	11,31351***	0,33	13,57581***	0,49	10,1009***	0,81
Lê livros	2,43179***	0,37	1,389692**	0,57	-1,741044***	0,69
Pais ou mãe tem ensino superior	1,502189***	0,55	6,217886***	0,79	24,39723***	0,99
Começou na creche	4,623788***	0,48	1,769434***	0,70	9,062857***	1,15
Começou na pré-escola	9,563907***	0,45	8,095357***	0,58	11,85789***	1,13
Aluno é Branco	3,205459***	0,38	6,288779***	0,60	8,621585***	0,80
Já foi reprovado	-16,34707***	0,39	-18,30068***	0,52	-30,15434***	0,95
Mora com 4 ou mais pessoas em casa	-11,07519***	0,30	-8,335363***	0,58	-4,135557***	0,77
Faz trabalho doméstico	-6,841176***	0,32	-0,2368046	0,50	-14,00325***	0,80
Tem computador com internet em casa	8,263845***	0,37	7,372987***	0,49	16,00738***	0,80
Norte	-12,70492***	0,60	-8,835195***	1,00	-12,14175***	2,38
Nordeste	-16,86206***	0,54	-9,937577***	0,72	-9,51936***	2,56
Sul	2,881263***	0,63	8,185768***	0,74	10,76021***	2,03
CentroOeste	-3,526546***	0,71	-0,5477002	1,03	-1,915818	2,24
Constante	190,2438***	0,62	219,0028***	0,86	247,5566***	2,63

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 5- Regressão Quantílica (4º decil)

	Quinta Série		Nona Série		Terceiro Ano	
	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão
Infraestrutura	8,182387***	0,52	6,549233***	0,72	21,11289***	0,53
Trabalha fora de casa	-11,28542***	0,49	-0,5997658***	0,56	-8,24979***	0,82
Frequenta biblioteca	-7,238086***	0,38	-5,233434***	0,66	-	-
Faz dever de casa de matemática	12,44963***	0,37	13,67807***	0,48	11,31043***	0,74
Lê livros	2,274569***	0,35	0,9150347**	0,51	-1,393836**	0,61
Pais ou mãe tem ensino superior	1,575467***	0,56	7,233202***	0,80	25,17598***	0,89
Começou na creche	5,148375***	0,54	1,714881***	0,67	9,058037***	1,12
Começou na pré-escola	10,06762***	0,44	7,816356***	0,57	11,14317***	1,21
Aluno é Branco	3,841931***	0,36	6,677436***	0,54	8,859135***	0,78
Já foi reprovado	-18,34622***	0,34	-18,28308***	0,44	-30,17106***	0,94
Mora com 4 ou mais pessoas em casa	-11,4988***	0,37	-8,602067***	0,52	-4,292903***	0,89
Faz trabalho doméstico	-6,960092***	0,35	-0,290515	0,44	-14,69273***	0,73
Tem computador com internet em casa	8,659866***	0,32	8,207858***	0,49	18,5141***	0,83
Norte	-14,30689***	0,64	-8,163793***	1,09	-9,293761***	2,54
Nordeste	-18,64053***	0,53	-9,9907***	0,64	-7,078697***	2,83
Sul	2,588381***	0,60	7,429493***	0,72	10,87741***	2,36
CentroOeste	-3,52964***	0,66	-0,818407	0,76	-0,681218	2,56
Constante	201,4524***	0,68	230,3783***	0,90	257,6859***	2,98

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 6- Regressão Quantílica (5º decil)

	Quinta Série		Nona Série		Terceiro Ano	
	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão
Infraestrutura	8,434089***	0,53	6,700408***	0,67	21,02923***	0,49
Trabalha fora de casa	-11,3794***	0,50	-0,399863	0,67	-8,466233***	0,91
Frequenta biblioteca	-7,786185***	0,43	-4,558015***	0,72	-	-
Faz dever de casa de matemática	13,102***	0,32	14,26102***	0,36	11,72901***	0,68
Lê livros	2,314483***	0,36	0,812822	0,56	-1,885122***	0,67
Pais ou mãe tem ensino superior	1,648333***	0,50	7,852023***	0,64	25,44466***	0,84
Começou na creche	5,559466***	0,50	2,015902***	0,65	7,882565***	1,07
Começou na pré-escola	10,93664***	0,40	7,92712***	0,57	9,252519***	0,98
Aluno é Branco	3,82676***	0,40	6,370243***	0,53	8,324048***	0,79
Já foi reprovado	-19,67038***	0,39	-18,59524***	0,45	-30,43029***	0,93
Mora com 4 ou mais pessoas em casa	-11,83968***	0,35	-9,065836***	0,47	-4,012696***	0,84
Faz trabalho doméstico	-6,978795***	0,37	-0,7457455*	0,44	-15,42968***	0,66
Tem computador com internet em casa	8,811406***	0,34	8,210169***	0,54	19,71779***	0,70
Norte	-16,09977***	0,78	-8,964913***	0,94	-9,03887***	2,12
Nordeste	-20,66176***	0,61	-10,43707***	0,62	-6,086662***	2,36
Sul	2,14004***	0,60	6,218418***	0,66	10,44199***	1,69
CentroOeste	-4,634241***	0,74	-0,387846	0,71	-0,61	1,99
Constante	212,5917***	0,59	241,6777***	0,91	271,213***	2,42

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 7- Regressão Quantílica (6º decil)

q60	Quinta Série		Nona Série		Terceiro Ano	
	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão
Infraestrutura	9,733814***	0,53	6,537913***	0,63	20,57357***	0,46
Trabalha fora de casa	-11,87501***	0,57	0,348500	0,62	-8,323455***	0,86
Frequenta biblioteca	-7,384647***	0,41	-4,344453***	0,69	-	-
Faz dever de casa de matemática	13,85942***	0,35	14,7265***	0,43	12,26217***	0,75
Lê livros	2,180868***	0,38	1,080306**	0,54	-2,03288***	0,64
Pais ou mãe tem ensino superior	1,985438***	0,49	8,512265***	0,78	25,27591***	0,98
Começou na creche	5,702791***	0,49	2,05737***	0,68	9,534452***	0,98
Começou na pré-escola	11,17528***	0,47	8,084088***	0,52	9,993676***	0,99
Aluno é Branco	4,150066***	0,36	6,004567***	0,51	8,447331***	0,78
Já foi reprovado	-21,06478***	0,37	-18,82338***	0,43	-28,97761***	0,92
Mora com 4 ou mais pessoas em casa	-12,44667***	0,43	-9,105536***	0,46	-3,740185***	0,84
Faz trabalho doméstico	-6,70213***	0,34	-1,237081***	0,44	-15,73963***	0,62
Tem computador com internet em casa	8,978064***	0,38	8,121448***	0,51	19,92541***	0,86
Norte	-16,76013***	0,67	-9,357321***	0,88	-10,2942***	1,96
Nordeste	-21,07054***	0,58	-11,04119***	0,75	-6,556485***	2,05
Sul	1,663343***	0,50	5,401243***	0,58	8,730074***	1,58
CentroOeste	-4,663361***	0,63	-0,306995	0,78	-1,236292	2,09
Constante	223,3468***	0,65	252,7785***	0,71	283,396***	2,06

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 8- Regressão Quantílica (7º decil)

	Quinta Série		Nona Série		Terceiro Ano	
	Coefficiente	Erro Padrão	Coefficiente	Erro Padrão	Coefficiente	Erro Padrão
Infraestrutura	10,38548***	0,45	6,808192***	0,59	20,33686***	0,49
Trabalha fora de casa	-12,16565***	0,57	0,635064	0,64	-8,095264***	1,02
Frequenta biblioteca	-7,223042***	0,43	-4,653647***	0,73	-	-
Faz dever de casa de matemática	15,06963***	0,37	15,32427***	0,42	12,42831***	0,74
Lê livros	1,998063***	0,41	1,400147**	0,59	-2,341187***	0,75
Pais ou mãe tem ensino superior	2,405491***	0,45	9,048759***	0,97	25,10434***	0,95
Começou na creche	5,924917***	0,54	2,263042***	0,72	8,185206***	1,05
Começou na pré-escola	11,29085***	0,45	7,637579***	0,59	9,350172***	1,07
Aluno é Branco	4,805707***	0,41	5,515507***	0,65	7,995777***	0,84
Já foi reprovado	-22,06165***	0,41	-18,70527***	0,49	-29,18884***	0,85
Mora com 4 ou mais pessoas em casa	-12,21261***	0,48	-9,040342***	0,53	-3,805047***	0,82
Faz trabalho doméstico	-6,321264***	0,40	-1,640243***	0,43	-15,69146***	0,79
Tem computador com internet em casa	9,381459***	0,42	7,921784***	0,58	20,5509***	0,94
Norte	-17,57588***	0,72	-9,825879***	1,17	-9,816834***	2,17
Nordeste	-22,11913***	0,57	-11,63647***	0,84	-6,319372***	2,16
Sul	0,7945655*	0,46	4,469507***	0,69	6,72349***	1,54
CentroOeste	-5,568856***	0,73	-0,916745	0,79	-1,941698	1,94
Constante	234,1353***	0,68	264,725***	0,82	297,7281***	2,41

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 9- Regressão Quantílica (8º decil)

	Quinta Série		Nona Série		Terceiro Ano	
	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão	Coeficiente	Erro Padrão
Infraestrutura	11,53693***	0,68	6,656036***	0,68	19,34163***	0,56
Trabalha fora de casa	-11,57324***	0,68	1,082576	0,82	-7,749882***	1,02
Frequenta biblioteca	-7,545182***	0,63	-4,749196***	0,89	-	-
Faz dever de casa de matemática	15,84993***	0,49	15,11974***	0,46	12,12245***	0,78
Lê livros	2,289879***	0,46	2,465164***	0,63	-3,075999***	0,64
Pais ou mãe tem ensino superior	2,491088***	0,68	9,519174***	0,92	23,07221***	0,98
Começou na creche	6,371471***	0,60	2,233704***	0,83	9,692522***	1,23
Começou na pré-escola	11,51092***	0,56	6,373622***	0,70	10,73389***	1,16
Aluno é Branco	5,057784***	0,46	5,985651***	0,77	7,138258***	0,96
Já foi reprovado	-23,90368***	0,53	-19,42754***	0,57	-28,92534***	0,92
Mora com 4 ou mais pessoas em casa	-12,3422***	0,45	-8,356162***	0,50	-2,555205***	0,74
Faz trabalho doméstico	-6,192442***	0,49	-1,862564***	0,49	-16,46245***	0,68
Tem computador com internet em casa	8,876996***	0,48	8,149599***	0,58	20,70756***	1,16
Norte	-17,59601***	0,71	-8,856429***	1,26	-8,351178***	2,55
Nordeste	-23,00939***	0,71	-11,29***	0,73	-4,349632*	2,37
Sul	-0,660303	0,64	4,023961***	0,71	7,050641***	2,16
CentroOeste	-6,178393***	0,75	-1,887797*	1,02	-1,597478	2,26
Constante	248,4624***	0,84	278,1525***	0,87	311,2411***	2,60

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 10- Regressão Quantílica (9º decil)

	Quinta Série		Nona Série		Terceiro Ano	
	Coefficiente	Erro Padrão	Coefficiente	Erro Padrão	Coefficiente	Erro Padrão
Infraestrutura	12,30906***	0,85	5,737583***	0,95	17,97468***	0,60
Trabalha fora de casa	-10,75899***	0,91	0,527592	0,89	-6,891062***	1,21
Frequenta biblioteca	-7,583085***	0,64	-3,274986***	1,30	-	-
Faz dever de casa de matemática	16,35179***	0,53	15,94002***	0,57	12,7942***	0,72
Lê livros	2,802234***	0,54	2,785091***	0,94	-2,786266***	0,67
Pais ou mãe tem ensino superior	2,543222***	0,82	10,32236***	0,87	21,14943***	1,06
Começou na creche	6,341906***	0,72	2,568205**	1,05	9,645701***	1,36
Começou na pré-escola	11,12768***	0,71	6,824243***	0,91	11,53081***	1,39
Aluno é Branco	4,752259***	0,58	6,049451***	0,67	6,713102***	0,92
Já foi reprovado	-24,8968***	0,57	-19,65441***	0,71	-27,95408***	1,16
Mora com 4 ou mais pessoas em casa	-12,90002***	0,57	-7,957884***	0,65	-2,376551***	0,83
Faz trabalho doméstico	-5,422028***	0,55	-2,392468***	0,61	-15,10992***	0,84
Tem computador com internet em casa	8,387729***	0,64	8,133084***	0,75	22,04153***	1,12
Norte	-17,63318***	0,87	-10,04582***	1,43	-4,047216**	2,08
Nordeste	-22,39219***	0,69	-11,03571***	0,79	-1,708025	2,16
Sul	-0,849765	0,88	2,588384***	0,93	8,494848***	1,85
CentroOeste	-7,632786***	1,12	-0,775386	1,17	-0,481732	1,82
Constante	268,473***	0,94	295,8208***	1,09	326,4107***	2,30

Obs: Erros-padrão robustos entre parênteses . * p<0.10, ** p<0.05, *** p<0.01.

Apêndice 11- Fatores de Inflação da Variância (VIF)

Variável	VIF		
	Quinta Série	Nona Série	Terceiro Ano
Infraestrutura	1,230	1,131	1,252
Branco	1,059	1,106	1,215
Tem computador com internet em casa	1,165	1,207	1,420
Mora com 4 ou mais pessoas	1,036	1,033	1,051
Lê livros	1,071	1,142	1,047
Frequenta bibliotecas	1,063	1,121	
Trabalha em casa 2 horas ou mais	1,010	1,007	1,083
Trabalha fora de casa	1,036	1,037	1,088
Já foi aprovado	1,107	1,096	1,088
Faz de verdade matemática	1,052	1,049	1,032
Creche	1,596	1,814	2,148
Pré-escola	1,592	1,804	2,050
Paio mãe tem ensino superior	1,061	1,060	1,287
Norte	1,222	1,148	1,502
Nordeste	1,491	1,437	1,727
Sul	1,121	1,149	1,326
Centro-Oeste	1,069	1,060	1,236

Obs: Vif > 10 significa problema de multicolinearidade