

UNIVERSIDADE FEDERAL DA PARAÍBA - UFPB
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA - PPGE
CURSO DE DOUTORADO EM ECONOMIA - CME

TRÊS ENSAIOS EM ECONOMIA SOCIAL

Linha de Pesquisa: Economia Regional e Políticas Públicas

FERNANDA LEITE SANTANA

JOÃO PESSOA – PB
2016

FERNANDA LEITE SANTANA

TRÊS ENSAIOS EM ECONOMIA SOCIAL

Tese apresentada ao Programa de Pós-Graduação em Economia (PPGE) da Universidade Federal da Paraíba (UFPB), como requisito para obtenção do título de Doutor(a) em Economia.

Orientador: Professor Pós-Dr. Erik Alencar de Figueiredo

**JOÃO PESSOA – PB
2016**

S232t Santana, Fernanda Leite.
Três ensaios em economia social / Fernanda Leite
Santana.- João Pessoa, 2016.
85f.
Orientador: Erik Alencar de Figueiredo
Tese (Doutorado) - UFPB/CCSA
1. Economia social. 2. Regressão quantílica - efeitos fixos.
3. Precipitação. 4. Desempenho acadêmico - tamanho - turma.
5. Mortalidade infantil.

UFPB/BC

CDU: 330.342.146(043)

Folha de Aprovação

**Aquilo que vivemos no sonho, e que nele vivemos repetidas
vezes, termina por pertencer à economia global de nossa
alma, tanto quanto algo “realmente” vivido.**

Nietzsche

AGRADECIMENTOS

Parece clichê, mas não tenho como fugir da afirmação de que a elaboração deste trabalho compreendeu muito mais que a titulação de doutora, pois foram envolvidas inúmeras pessoas e instituições, sem as quais eu não poderia estar realizando esse sonho.

Certo dia, nos meus anseios por escritas inspiradoras dessa jornada, li um texto que fala do terrível custo de se fazer um doutorado, abordando as sequelas e os efeitos psicológicos adquiridos ao longo do processo, mas que em nenhum momento o vi abordando o apoio e as amizades que se conquistam nesse tempo.

Nesse sentido, parafraseando os conteúdos dos livros de econometria que me deparei ao longo do curso, posso afirmar que apesar do doutorado ser um resultado robusto, não tem como estimar a significância das pessoas na determinação do mesmo como se fosse uma simples regressão, pois aprendi que relacionamentos não são determinísticos.

E, por mais dolorido que possa ter sido, o que fica é uma imensa gratidão como reconhecimento de que, por mais que um trabalho acadêmico possa parecer individual, ele é na sua essência fruto de muitos esforços individuais e coletivos. E nessa conjuntura, externo aqui meus agradecimentos:

À Deus, por fazer parte da minha vida em todos os momentos.

À minha família, meus pais Margarida e Marinaldo pela cooperação e paciência durante todo o curso e por acreditar no meu potencial para concluir essa etapa. Aos meus irmãos André Luiz e Thyago pelo apoio.

À família de Jairo, que também posso chamar de minha família, vivenciaram comigo momentos de alegria e amor fraterno que jamais poderei esquecer.

Aos amigos acadêmicos que os levarei no coração para o resto da minha vida, Patrícia Soares, Etevaldo Almeida, Ana Annegues, Wallace Patrick, Herbert Vinícius, Paola Freire e Herbert Rêgo. Foram momentos inesquecíveis.

À Erik Alencar, orientador, obrigado pela compreensão e paciência nos momentos difíceis que me deparei ao longo deste trabalho. Sua prontidão e disponibilidade *online* foram essências para mim. Meu muito obrigado!

Aos pesquisadores: Francisco Soares de Lima, Jorge Luiz Mariano da Silva, José Luís da Silva Netto Júnior e Shirley Pereira de Mesquita, pela disponibilidade de participar da avaliação deste trabalho.

Aos funcionários do PPGE-UFPB, Risomar Farias e Ricardo, pela presteza e amabilidade nas demandas requeridas.

À Conselheira-Presidente do Tribunal de Contas do Estado do Acre – TCE-AC, *Naluh Maria Lima Gouveia*, pela oportunidade concedida em encerrar esta etapa da minha vida. Sou eternamente grata por sua sensibilidade a minha demanda.

Aos amigos que fiz no TCE-AC, em especial a minha chefe Semírames Maria Plácido Dias, por acreditar em mim, como também pelo apoio da Laelia, Kelly e Marnízia e demais amigos da 2ª Inspeção.

Agradecimento Especial

Ao meu esposo Jairo, amigo de todas as horas com o qual pude compartilhar sonhos e conquistas ao longo dos 14 anos de estrada que temos. Sem a sua reciprocidade e companheirismo eu não chegaria a tal ponto.

RESUMO

Este trabalho é composto por três artigos na área de Economia Social, que se dedicam a estudar três identificações empíricas sob o escopo de estimadores com controles individuais para efeitos individuais. Para tanto, constroem-se painéis quantílicos para cada desenho empírico conforme a literatura existente e o arcabouço teórico. No primeiro ensaio investigam-se os efeitos do tamanho da turma no desempenho acadêmico dos universitários que ingressaram na UFPB no período de 2001 à 2010. Para estimadores condicionais a mediana, os resultados sugerem uma relação negativa entre o tamanho da classe e o desempenho acadêmico, levando a concluir que políticas de redução de turma teria um efeito benéfico no desempenho acadêmico dos alunos da UFPB. Contudo, ao utilizar o estimador QR(FE), averigua-se um valor superestimado desse efeito decorrente do viés do estimador padrão por não captar a heterogeneidade não observada. Além disso, a análise ao longo da distribuição condicional de desempenho evidencia que existe heterogeneidade de efeitos. Os resultados aqui evidenciados além de trazer novos elementos para as políticas públicas, também abrem uma agenda de pesquisa relevante em virtude da falta de consenso na literatura do efeito da variável de interesse. Por sua vez, o segundo ensaio traz a luz questões sobre os determinantes do desempenho acadêmico dos universitários ao analisar a relação causal entre a disponibilidade de água pluvial no período gestacional dos indivíduos nascidos no semiárido brasileiro e o resultado educacional individual. O pressuposto é de que as chuvas em ano de nascimento tem relação positiva com o desenvolvimento das habilidades cognitivas e não cognitivas dos indivíduos quando adultos. Neste cenário, análogo a análise anterior, o estimador condicional à mediana superestima o efeito da variável de interesse, à qual consiste no log-desvio das chuvas 12 meses antes do nascimento do indivíduo. No geral, tem-se que o efeito é positivo, bem como persistente ao longo da distribuição condicional de desempenho acadêmico. Tais constatações são relevantes às políticas públicas, pois evidenciam que investimentos em idades precoces têm efeito maior do que investimentos feitos mais tarde, o que levanta questões sobre o papel de políticas de acesso ao ensino superior, sistema de cotas, entre outros. Por fim, o terceiro ensaio contribui para as políticas públicas ao ressaltar a importância das questões climáticas nos resultados de saúde. Para avaliar os efeitos das chuvas, realizaram-se três exercícios empíricos: i) estimador de regressão quantílica com efeitos fixos ii) estimador de regressão quantílica com efeitos fixos e variáveis instrumentais; iii) dados de precipitação observacional. Os resultados sugerem que: i) quanto maior a precipitação antes da concepção, durante o período gestacional e nos primeiros anos de vida dos indivíduos menor é a taxa de mortalidade infantil nesta região; ii) quando controlada a heterogeneidade não observada e a endogeneidade existente na relação, a precipitação tem menor peso indicando que métodos condicionais à média superestimam essa relação. Em síntese, os resultados corroboram com a literatura de que os bebês nascidos nas zonas áridas e que passam por períodos de seca quando no útero de suas mães enfrentam risco maior de morte, especialmente quando as secas comprometem os períodos antes e após a concepção dos indivíduos.

Palavras-chaves: Economia Social; Regressão Quantílica com Efeitos Fixos; Tamanho da Turma; Precipitação; Desempenho Acadêmico; Mortalidade Infantil.

ABSTRACT

This work consists of three articles in the area of social economics, which are dedicated to study three empirical identifications under the scope of estimators with individual controls for individual effects. For this purpose, are constructed quantile panels for each empirical design as the existing literature and theoretical framework. In the first trial we investigate the effects of class size on academic performance of students who entered the UFPB from 2001 to 2010. For conditional estimates of the median, the results suggest a negative relationship between class size and academic performance, leading to the conclusion that class reduction policies would have a beneficial effect on the academic performance of UFPB students. However, when using the estimator QR (FE), we find out a overestimated value of this effect resulting from the standard estimator bias, because it does not capture the unobserved heterogeneity. Furthermore, analysis along the conditional distribution performance demonstrates that there is heterogeneity effects. The results highlighted here bring new elements to public policy and also open a relevant research agenda because of the lack of consensus in the literature of the effect of the variable of interest. In turn, the second trial brings to light questions about the determinants of academic performance of students analyzing the causal relationship between the rain water availability during pregnancy of individuals born in the Brazilian semi-arid and individual educational outcome. The assumption is that the rains in year of birth has a positive relationship with the development of cognitive and non-cognitive skills of individuals as adults. In this scenario, similar to the previous analysis, the conditional median estimator overestimate the effect of the variable of interest, which is the log deviation of rain 12 months before the birth of the individual. In general, the effect is positive, as well as persists along the conditional distribution of academic achievement. Such findings are relevant for public policies as evidence that investment in early childhood has a greater effect than investments done later, which raises questions about the role of access policies to higher education, as the quota system, among others. Finally, the third test contributes to public policies to emphasize the importance of climate issues in health outcomes. To assess the effects of rain, we make three empirical exercises: i) quantile regression estimator with fixed effects ii) quantile regression estimator with fixed effects and instrumental variables; iii) observational precipitation data. The results suggest that: i) the higher the rainfall before conception, during pregnancy and the first years of life of individuals, the lowest infant mortality rate in the region; ii) when controlling the unobserved heterogeneity and endogeneity in the relationship, the rainfall has less weight indicating that methods conditional to mean overestimate this relationship. In summary, the results corroborate the literature that babies born in arid and go through periods of drought when in the womb of their mothers face greater risk of death, especially when drought undertake the periods before and after conception of individuals .

Keywords: Social Economics; Quantile regression with fixed effects; Size of the Group; Precipitation; Academic performance; Child mortality.

LISTA DE TABELAS

ENSAIO 1

Tabela 1.1 -	Perfil do aluno da UFPB de 2001 à 2010	18
Tabela 1.2 -	Número de docentes por Centro Acadêmico e por Titulação	19
Tabela 1.3 -	Estatísticas Descritivas Geral	22
Tabela 1.4 -	Estatística Descritiva por Centro Acadêmico – Média	23
Tabela 1.5 -	Resultados Empíricos para a mediana	26

ENSAIO 2

Tabela 2.1 -	Estatística Descritiva	50
Tabela 2.2 -	Resultados das estimações	52

ENSAIO 3

Tabela 3.1 -	Correlação de <i>Spearman</i> entre dados de Precipitação de TSM	71
Tabela 3.2 -	Resultados das estimações com método de Canay (2011) do período de 1996-2010 da Região Semiárida do Nordeste	73
Tabela 3.3 -	Resultados das estimações com método de Harding e Lamarche (2009) do período de 1996-2010 da Região Semiárida do Nordeste	76
Tabela 3.4 -	Resultados das estimações de comparação de dados de Reanálise e Observacionais do período de 1996-2010 da Região Semiárida do Nordeste	78

LISTA DE GRÁFICOS

ENSAIO 1

Gráfico 1.1	Estimativas do impacto do tamanho da classe no desempenho acadêmico dos indivíduos ao longo da distribuição	27
Gráfico 1.2	Estimativas ao longo da distribuição por centro acadêmico – CCEN, CCTA e CCS	28
Gráfico 1.3	Estimativas ao longo da distribuição por centro acadêmico – CCSA, CCHLA	29
Gráfico 1.4	Efeitos quantílicos das covariáveis sobre o desempenho acadêmico	30

ENSAIO 2

Gráfico 2.1	Evolução por <i>quantis</i> dos coeficientes estimados por QR(FE)	53
--------------------	---	----

ENSAIO 3

Gráfico 3.1	Estimativas do impacto das secas sob os resultados dos indivíduos ao longo da distribuição	75
--------------------	--	----

LISTA DE QUADROS

ENSAIO 2

Quadro 2.1	Descrição das Variáveis	48
-------------------	-------------------------------	----

ENSAIO 3

Quadro 3.1	Diversas Fontes e Metodologias de dados de Precipitação.....	68
-------------------	---	----

SUMÁRIO

INTRODUÇÃO	13
 CAPÍTULO 1: O Impacto do Tamanho da Classe no Desempenho	 14
Acadêmico Superior: Efeitos Condicionais Heterogêneos	
Resumo -----	14
Abstract -----	14
1.1 Introdução -----	15
1.2 Contexto Institucional e Perfil do Aluno -----	17
1.3 Estratégia Empírica e Dados -----	21
1.3.1 <i>Estratégia empírica</i> -----	21
1.3.2 <i>Descrição dos Dados</i> -----	23
1.4 Resultado e Discussões -----	26
1.4.1 <i>Linha Base</i> -----	26
1.4.2 <i>Política de Redução do Tamanho da turma por Centro Acadêmico</i> ---	29
1.4.3 <i>Efeitos das Covariáveis</i> -----	31
1.5 Considerações Finais -----	33
1.6 Referências Bibliográficas -----	34
Apêndice -----	36
Anexo -----	37
 CAPÍTULO 2: Escassez de água e os Resultados Educacionais dos	 38
Universitários do Semiárido Brasileiro	
Resumo -----	38
Abstract -----	38
2.1 Introdução -----	39
2.2 Formação de Capacidades -----	41
2.1 <i>Formação de Habilidades: Psicologia, Economia, Neurociência e</i>	
<i>Saúde</i> -----	41
2.2 Estratégia de Identificação Empírica -----	45
2.2.1 <i>Modelo Teórico</i> -----	45
2.2.2 <i>Identificação Empírica</i> -----	46

2.2.3 <i>Descrição dos Dados</i> -----	48
2.3 Resultados -----	51
2.3.1 <i>Estatística Descritivas</i> -----	51
2.3.2 <i>Efeitos das secas no desempenho acadêmico</i> -----	52
2.4 Considerações Finais -----	56
2.5 Referências Bibliográficas -----	57

CAPÍTULO 3: Escassez da Água e Mortalidade Infantil: Novas Evidências a partir de um Painel Quantílico Robusto -----	60
Resumo -----	60
Abstract -----	60
3.1 Introdução -----	61
3.2 Metodologia e Dados -----	64
3.2.1 <i>Contexto</i> -----	64
3.2.2 <i>Estratégia Empírica</i> -----	65
3.2.3 <i>Descrição dos Dados</i> -----	68
3.3 Resultados Empíricos -----	73
3.3.1 <i>Heterogeneidade não observada</i> -----	73
3.3.2 <i>Endogeneidade</i> -----	77
3.3.2 <i>Dados Observacionais</i> -----	78
2.4 Considerações Finais -----	81
2.5 Referências Bibliográficas -----	82
Anexo -----	85

CONSIDERAÇÕES FINAIS	86
-----------------------------------	----

INTRODUÇÃO

O objetivo da tese concentra-se em utilizar técnicas de identificação econométricas, em especial, estimadores com controles individuais para efeitos individuais, focando em temas de economia social sob o ponto de vista das políticas públicas. Especificamente, busca-se adicionar elementos à literatura da economia da educação no primeiro e segundo ensaio e à economia da saúde no terceiro ensaio, fornecendo *insights* sobre a distribuição condicional dos resultados econômicos em educação e saúde, respectivamente.

O primeiro ensaio tem-se como propósito averiguar os efeitos do tamanho da turma no desempenho acadêmico dos indivíduos. Em cenário onde a literatura sobre o tema é limitada e ainda não explorada para o ensino superior do Brasil, este trabalho acrescenta a utilização de regressão quantílica para uma estrutura de dados em painel apreciado em Canay (2011). Além disso, utiliza-se de uma base única de registros acadêmicos da Universidade Federal da Paraíba (UFPB) para a construção de um painel quantílico *aluno-turma-período* que incluem estudantes entre 2001 e 2010 na busca de evidências sobre a relação causal do tamanho da turma na função de produção educacional.

Já o segundo ensaio dedica-se a investigar as relações entre clima e resultado educacional. Parte-se do pressuposto de que o indivíduo recebe influências genéticas, ambientais, investimento ao longo da vida e herdabilidade parental que determinam a formação do seu estoque de capacidades. Logo, embasado pelo modelo teórico de Cunha e Heckman (2007) para a tecnologia de formação de habilidades, utiliza-se a abordagem metodológica de Canay (2011) em um painel quantílico *aluno-disciplina-ano* dos discentes da UFPB que nasceram no semiárido brasileiro. Em linhas gerais, procura-se inferir sobre a variável de interesse precipitação nos últimos 12 meses antes do nascimento dos indivíduos sobre o desempenho acadêmico na faculdade, expresso nas notas das disciplinas cursadas ao longo dos cursos de graduação, controladas por variáveis socioeconômicas e *background* familiar.

Por fim, o terceiro ensaio avalia as implicações de restrições pluviométricas sobre os períodos iniciais de vida dos indivíduos e seus efeitos nos resultados de saúde expressos na taxa de mortalidade infantil. A partir de um painel *município-a-mês* da região semiárida brasileira, adota-se uma estratégia empírica que envolve três objetivos específicos. Em primeiro lugar, aplica-se a metodologia baseada em Canay (2011) para captar a heterogeneidade não observada e os efeitos das covariáveis em ambientes de efeito fixo. Segundo, visando endogeneizar a quantidade de chuvas, utiliza-se a Temperatura da Superfície do Mar (TSM) como instrumento em um modelo empírico baseado em Harding e Larmarche (2009). E em terceiro lugar, verifica-se a confiabilidade dos dados de reanálise a partir de dados observacionais para precipitação.

A estrutura da tese é dividida além desta introdução, em três capítulos dedicados a apresentar em cada um os referidos ensaios aqui expostos. Ao final apresentam-se as considerações finais do estudo.

CAPÍTULO 1: O Impacto do Tamanho da Classe no Desempenho Acadêmico Superior: Efeitos Condicionais Heterogêneos

Resumo

O presente ensaio avalia o impacto causal do tamanho da classe no desempenho acadêmico dos alunos universitários. Para isso, utiliza-se uma base de dados única de registros administrativos no período de 2001 a 2010, sobre o desempenho acadêmico dos discentes da Universidade Federal da Paraíba (UFPB), bem como informações sobre as características individuais, *background* familiar e a nota de entrada do vestibular. Além disso, aplica-se o método de regressão quantílica em estruturas de dados em painel abordado em Canay (2011) para um painel *aluno-turma-período*. No que concerne os resultados obtidos, observou-se que as estimações são estatisticamente significativas ao longo da distribuição, com efeitos negativos para cauda inferior e efeito positivo para a cauda superior. Além disso, as análises indicam que estudantes adicionais do sexo feminino ou aqueles com melhores resultados nos testes de entrada tendem a melhorar o desempenho da turma em geral, e, por outro lado, mais aluno com renda alta tem efeito negativo, sobretudo na cauda inferior.

Palavras-chaves: Tamanho da Turma; Regressão Quantílica com efeitos fixos; Desempenho acadêmico; Políticas de Redução de turmas.

Abstract

This paper evaluates the impacts of class size on academic performance of college students. For this, it uses a single database of administrative records in the 2001-2010 period, about the academic performance of students of the Federal University of Paraíba (UFPB) as well as information on individual characteristics, family background and note Vestibular input. Moreover, it applies the quantile regression method discussed panel data structures in Canay (2011) for a student-class time-panel. Concerning the results obtained, it was observed that the estimates are statistically significant throughout the distribution, with negative effects on the bottom tail and a positive effect on the upper tail. Furthermore, analyzes indicate that additional female students or those with better results at the entry tests tend to improve the performance of the class in general and, on the other hand, more student with high income has a negative effect, especially in the lower tail .

Keywords: Class Size; Quantile regression with fixed effects; Academic performance; Reduction Policies classes.

1.1 INTRODUÇÃO

Em toda a literatura que aborda questões referentes à educação, ainda não há consenso de quais/como são os principais *inputs* determinantes da função de produção educacional¹ e seus efeitos. Um dos exemplos claros de controvérsia é o efeito do tamanho das turmas. Para o ensino básico², alguns estudos estimam que o tamanho da turma possui impactos negativos (Angrist e Lavy, 1999; Krueger, 1999; Urquiola, 2006), misto (Duflo et. al., 2000; Alam, 2000) ou até mesmo nenhum impacto sobre os resultados educacionais (Hoxby, 2000; Asadullah, 2005; Banerjee et. al., 2005). Esses resultados contrastantes podem estar relacionados a graves problemas econométricos que refletem na dificuldade de mensurar os verdadeiros efeitos causais dessa relação.

Quanto ao ensino universitário, apesar de uma literatura incipiente, as poucas evidências já são perceptíveis de que também não existe um consenso. Por exemplo, Machado e Vera-Hernandez (2008), a partir dos dados administrativos da Universidade Carlos III de Madri na Espanha, não encontraram qualquer efeito significativo do tamanho da classe no desempenho dos alunos. De Giorgi, et. al., (2012), avaliando a Universidade de Bocconi na Itália, descobrem que o tamanho das turmas tem um impacto pequeno, mas negativo e significativo. Efeitos negativos também são encontrados em Bedard e Kuhn (2006) e Monks e Schmidt (2011), para a Universidade da Califórnia nos EUA e uma Universidade Privada da costa leste dos EUA, respectivamente. Já Bandiera, Larcinese e Rasul (2010), investigaram Universidade líder do Reino Unido e concluíram que existe um efeito negativo na média, contudo, para determinados tamanhos de classes tem pouco efeito sobre os resultados dos testes. Mais recentemente, De Paola, Ponzio e Scoppa (2013), usando dados da Universidade da Calábria, no interior da Itália, constataram que classes maiores determinam um efeito negativo e significativo sobre o desempenho dos alunos, mas ao considerar uma *dummy* para classes grandes e pequenas, remataram que esse efeito negativo é significativamente maior para estudantes de baixa capacidade.

Regra geral, as estimativas costumam introduzir variáveis relacionadas aos indivíduos como o *background* familiar e outras características individuais, dados sobre professores, insumos escolares, entre outros, embora variem na configuração de diferentes estratégias de identificação, metodologia adotada e contexto institucional. Uma característica comum à maioria dos estudos sobre o efeito do tamanho da classe, sobretudo no ensino superior, é o foco na média da distribuição de desempenho acadêmico. Neste contexto, Harding e Lamarche (2014) adotam um estimador de regressão quantílica em dados em painel para variáveis endógenas, e avançam na literatura ao concluir que existe heterogeneidade de efeitos ao longo da distribuição condicional do desempenho acadêmico dos alunos. Os autores ressaltam que esses resultados são consistentes com as atuais pesquisas na área de psicologia educacional e relevante para os decisores de políticas públicas.

¹ A Função de Produção de Educação, denominada de FPE, analisa a forma como os diversos insumos do processo educacional podem afetar os resultados educacionais dos indivíduos (Hanushek, 1970, 1979).

² Chamadas de escolas primárias e secundárias na literatura internacional, que correspondem ao ensino fundamental e médio no Brasil.

Comparado a literatura internacional, a investigação sobre os efeitos do tamanho da classe do ensino básico no Brasil é muito mais limitada e consequentemente controversa, por exemplo, temos Menezes-Filho (2007), Oliveira e Menezes-Filho (2008), Camargo e Sabino Porto Júnior (2014)³, dos quais encontraram nenhum efeito, efeito negativo e efeito relativamente misto, respectivamente. Outrora, até o presente momento não se tem conhecimento de nenhum outro estudo voltado a analisar o papel do tamanho das turmas de faculdade no desempenho acadêmico dos alunos universitários do Brasil. Dessa forma, o presente artigo vem a contribuir com a literatura em cenário onde os cursos universitários do Brasil são um laboratório sub-explorados no que diz respeito aos efeitos do tamanho da classe.

A propósito, este artigo analisa os registros administrativos da Universidade Federal da Paraíba para identificar o impacto causal do tamanho na classe nas notas obtidas nas disciplinas ao longo dos cursos de graduação da instituição. Seu escopo é: *i)* mostrar evidências da importância da turma na explicação do desempenho distributivo; *ii)* estratificar a amostra por centro acadêmico a fim de estimar alguns efeitos do tamanho da turma e suas implicações para as políticas de redução de classe; e, por fim, *iii)* avaliar os efeitos das covariáveis estudadas.

Para alcançar esses objetivos, analisa-se a distribuição condicional de desempenho acadêmico por meio da aplicação do método de regressão quantílica com dados em painel abordado em Canay (2011). Essa abordagem leva em consideração os indícios da literatura de que os diferentes *quantis* respondem de forma distinta a alterações nas variáveis explicativas. O argumento para isso é a existência de heterogeneidade não observada entre os indivíduos e efeitos fixos de localização, como por exemplo, a motivação ou habilidade do aluno em absorver conhecimento e qualidade de ensino. Assim sendo, características não observáveis correlacionadas com o tamanho da turma e com as covariáveis poderiam ser responsáveis por uma parcela dos resultados educacionais. Ademais, utilizar-se-á a especificação de De Giorogi et. al (2012), exceto o uso da variação experimental⁴, uma vez que não existe na UFPB.

Além dessa introdução, o trabalho organiza-se em mais quatro seções. A seção 1.2 trata do contexto institucional abordando características da instituição, bem como traça o perfil do estudante a fim de captar informações associadas as variáveis de controle utilizados nas estimações. A seção 1.3 discute a estratégia empírica e descreve os dados utilizados aqui. A seção 1.4 destina-se os resultados sobre o desempenho acadêmico. Finalmente, a seção 1.5 as considerações finais.

³ Os autores investigam o efeito do acréscimo de alunos na sala de aula, não a redução das turmas.

⁴ Os autores utilizam dados baseado numa variação experimental que surge a partir de um mecanismo de atribuição aleatória de alunos para dar aula na Universidade de Bocconi.

1.2 CONTEXTO INSTITUCIONAL E PERFIL DO ALUNO

A análise aqui requerida é baseada nos dados administrativos sobre cada aluno da Universidade Federal da Paraíba (UFPB), para os anos de 2001 à 2010. A UFPB⁵ é a maior Instituição Pública de ensino superior existente no Estado. No total possui 72 cursos de graduação distribuídos em 13 (treze) Centros acadêmicos⁶ que compõem quatro Campi⁷ Universitários. O Campus I⁸, localizada na cidade de João Pessoa, é o maior campi e possui maior quantidade de matriculados, e, portanto, foco da análise desse estudo.

O ingresso de estudantes pode ocorrer de duas formas tradicionais: *i)* Vestibular no qual testa os conhecimentos do estudante nas disciplinas cursadas no ensino médio; *ii)* Exame Nacional de Ensino Médio (ENEM) que também tem questões relacionadas ao aprendizado do ensino médio e que somente passou a ser forma de ingresso obrigatória após 2010, e, com isso, não pertence a nossa análise. Além das formas tradicionais, ainda tem-se convênios, mudanças de campus, alunos graduados, transferência, reopção de curso, reingresso, entre outros.

O sistema de ensino comporta três fases: uma graduação, que pode ser licenciatura ou bacharelado, com duração média de 4 (quatro) à 5 (cinco) anos; um mestrado de duração de 2 (dois) anos e um doutorado em média 4 (quatro) anos. Contudo, para fins desta análise concentrar-se-á apenas em estudantes dos cursos de graduação que foram matriculados em disciplinas de 2001 à 2010, dos quais foram ofertados 5.047 disciplinas distintas. Para cada curso existe um departamento específico que coordena as atividades de ensino e extensão, bem como a sua logística.

Cada programa de graduação possui um currículo acadêmico com sua própria lista de disciplinas básicas (obrigatórias) ou oferecendo uma escolha de disciplinas eletivas, além do que estas podem incluir disciplinas pertencentes a outros departamentos acadêmicos. O período letivo é semestral, no qual o aluno se matricula nas disciplinas que pretendem cursar. Ao final de cada semestre, cada estudante deve obter uma nota igual ou superior a 7 para ser aprovado por média ou maior ou igual a 5 após os exames finais. Ao final do curso o aluno obtém o Coeficiente de Rendimento Escolar que nada mais é do que o somatório ponderado de todas as disciplinas cursadas ao longo do tempo e adicionado ao trabalho de conclusão de curso confere ao aluno o referido diploma de graduação.

Normalmente, o aluno que ingressa na Instituição automaticamente já é matriculado nas disciplinas que compõem o primeiro semestre da grade curricular pré-definida. No

⁵ Além da UFPB, tem-se a Universidade Estadual da Paraíba (UEPB); Instituto Federal da Paraíba (IFPB) e a Universidade Federal de Campina Grande (UFCG).

⁶ A lista dos cursos de graduação e seus respectivos Centros Acadêmicos encontram-se no Anexo I.

⁷ A estrutura da UFPB comporta: *i)* Campus I (na cidade de João Pessoa); *ii)* Campus II (Areia); *iii)* Campus III (Bananeiras); *iv)* Campus IV (Mamanguape e Rio Tinto).

⁸ Fazem parte os seguintes Centros: Centro de Ciências Humanas, Letras e Artes - CCHLA; Centro de Educação – CE; Centro de Comunicação, Turismo e Artes - CCTA; Centro de Ciências da Saúde - CCS; Centro de Tecnologia - CT e Centro de Ciências Jurídicas - CCJ; Centro de Ciências Exatas e da Natureza – CCEN; Centro de Biotecnologia - CBIotec; Centro de Tecnologia e Desenvolvimento Regional - CTDR; Centro de Ciências Sociais Aplicadas - CCSA ;Centro de Ciências Médicas - CCM; ; Centro de Informática - CI e Centro de Energias Alternativas Renováveis – CEAR

entanto, este aluno tem a opção de trancamentos, mesmo no primeiro semestre. Sem trancamentos, para cada semestre teríamos turmas de tamanhos iguais, que no universo grande, a probabilidade disso ocorrer é pequena. Além disso, com as mais diversas formas de ingresso nos cursos, bem como a existência de alunos que repetem, abandonam, transferem, fazem com que o tamanho da turma varie ao longo do tempo. Desta variação pressupõe-se que são correlacionados com o interesse e capacidade do aluno e aqui permite ser palco de debate político e acadêmico a respeito do seu impacto no desempenho educacional dos alunos.

No intuito de caracterizar a demanda estudantil da UFPB constatou-se que nos últimos anos a instituição vem implantando um conjunto de políticas de expansão que conferiram o aumento de matrículas. Para ilustrar, de 2005 para 2011 a instituição elevou o número de cursos de graduação de 50 para 104, bem como o número de alunos matriculados que passou de 18.759 para 29.629 estudantes (NTI, 2012), seja pelo aumento da oferta ou da demanda propriamente dita. Sabe-se que o aumento de matrículas importa em ajustes de *inputs*, como por exemplo, a contratação de novos professores, expansão de espaço físico, contratação de novos funcionários administrativos, entre outros, que muitas vezes necessitam de um tempo maior para sua implantação. Esse lapso temporal entre os ajustes pode trazer a tona questões como aumento do tamanho da classe para comportar essa nova demanda.

Além disso, as salas de aulas possuem capacidades físicas limitantes que comprometem o número de matriculado por disciplina. Hoje, há cerca de 70 departamentos gerenciando os mais diversos tamanhos de sala de aula. Não obstante, o aluno também se depara com problemas de cunho pessoal ou profissional que compromete a duração do curso, fato comprovado pelo número de egresso (após o período mínimo de curso) que difere quantitativamente do egresso⁹. Tais fatos evidenciam que o aluno se depara normalmente com turmas maiores ou menores ao longo do curso.

Neste cenário, investiga-se o impacto do tamanho da classe no desempenho acadêmico dos alunos que ingressaram nos cursos de graduação da UFPB, e este é possibilitado por algumas razões. Primeiro, a existência de uma gama de informações obtidas junto ao Núcleo de Tecnologia da Informação (NTI) dos quais se extraem informações concernentes ao desempenho acadêmico e características individuais dos alunos. Segundo, uma vez que de certa forma o aluno tem a opção de escolher a disciplina e o período em que irá cursá-la, há aleatoriedade na formação e composição das turmas. Terceiro, o uso das notas das disciplinas parecer ser uma *proxy* mais confiável para resultado educacional individual do que os testes padronizados comumente utilizados na literatura. A Tabela 1.1 traça o perfil do aluno aqui estudado.

A fim de capturar informações a respeito dos discentes do período de 2001 à 2010, a Tabela 1.1 sumariza as principais informações abrangendo características individuais. Verifica-se que cerca de 51% são mulheres; 55% são da cor branca e na grande maioria possui renda familiar baixa de até 3 salários mínimos. Em termos de escolaridade do pai, observa-se elevada taxa de pai sem instrução (24,57%) que somado aos pais de ensino

⁹ Ver Apêndice A1.

fundamental (14,34%) totaliza quase 40% da amostra, denotando um baixo *background* familiar. Apesar disso, boa parte da amostra vem de escola privada tanto no ensino fundamental (57,54%) quanto no ensino médio (56,28%), possuem computador em casa (71,28%) e no momento do ingresso na Instituição não exercia trabalho remunerado (74,44%).

Tabela 1.1 – Perfil do aluno da UFPB de 2001 à 2010

Sexo		Raça	
Feminino	Masculino	Brancos	Não Brancos
51.30%	48.70%	55.90%	44.10%
(532.55)	(544.69)	(545.02)	(527.79)
Renda Familiar			
Mínima	Baixa	Média	Alta
6.47%	46.56%	33.07%	13.90%
(493.89)	(510.19)	(547.81)	(579.80)
Instrução do Pai			
Analfabeto	Fundamental	Médio	Superior ou mais
24.57%	14.34%	33.26%	27.83%
(504.64)	(513.82)	(531.82)	(563.26)
Mãe trabalha?		Pai Trabalha?	
Sim	Não	Sim	Não
80.30%	19.70%	73.60%	26.40%
(535.76)	(517.95)	(539.74)	(517.01)
Tipo de Escola do Ensino Fundamental		Tipo de Escola do Ensino Médio	
Pública	Privada	Pública	Privada
42.46%	57.54%	43.72%	56.28%
(504.66)	(551.01)	(504.78)	(552.14)
Aluno exerce algum trabalho		Tem computador em casa	
Sim	Não	Sim	Não
25.56%	74.44%	71.28%	28.72%
(517.95)	(535.91)	(542.11)	(504.23)

Fonte: Elaboração Própria a partir dos dados fornecidos pelo NTI (UFPB). Os valores em % representam a participação percentual do total da amostra com informações disponíveis. Os Valores entre parênteses correspondem a média obtida no vestibular pelo grupo de indivíduos.

Outro ponto a ser destacado são os resultados dos testes de pré-inscrição. Ao calcular uma média para cada grupo – valores entre parênteses na Tabela 1.1 – confirmam-se alguns preceitos da literatura. Por exemplo, alunos de renda familiar baixa possuem nota média de entrada menor do que aqueles com renda mais alta. O mesmo ocorre para aqueles que têm pais analfabetos ou mesmo aqueles em que pai ou mãe que não trabalham. Destarte, estudantes do sexo masculino, brancos, com renda familiar alta, filhos de pais que trabalham, alunos oriundos de escola privada, tem computador em casa e não exerciam trabalho remunerado na época do teste de entrada tem um melhor desempenho.

No que tange o perfil dos docentes, a Tabela 1.2 apresenta os alguns números por centro acadêmico e titulação. Verifica-se que durante o período analisado (2001-2010), grande parte do corpo docente tem titulação de doutorado, o que significa cerca de 40% do

total. Note que boa parte dos doutores concentram-se nos centros CT e CCS, dos quais os cursos de engenharia, odontologia, nutrição, entre outros, se encontram.

Tabela 1.2 – Número de docentes por Centro Acadêmico e por Titulação

Centro Acadêmico/Titulação do Docente	Especialização	Graduação	Mestrado	Doutorado	Pós-doutorado	Total por Centro Acadêmico
CCEN	7	49	62	134	0	252
CCHLA	10	56	91	177	1	335
CCM	39	32	42	42	0	155
CE	4	16	48	53	0	121
CCSA	18	101	106	98	0	323
CT	13	72	96	196	0	377
CCS	35	132	167	187	0	521
CCJ	16	38	29	20	1	104
CCTA	13	101	81	64	1	260
CEAR	1	1	0	2	0	4
CI	0	7	8	9	0	24
Total por Titulação	156	605	730	982	3	2.476

Fonte: Elaboração Própria a partir dos dados do NTI (UFPB).

É possível, ainda, destacar dois pontos. Primeiro, considerando titulação como *proxy* para qualidade dos docentes, o que, por sua vez, seria indicativo da qualidade de ensino ofertada, poderia inferir *a priori* que a qualidade educacional da instituição é baixa tendo em vista que cerca de 60% do total da amostra tem no máximo o título de mestre. Segundo, reforçar-se tal constatação pelo inexpressivo número de pós-doutores no geral e por centro, como também a existência de docentes com apenas especialização e o elevado número de professores apenas com graduação.

1.3 ESTRATÉGIA EMPÍRICA E DADOS

Neste artigo busca-se estimar o efeito distributivo do tamanho da classe controlando as características individuais e *background* familiar dos discentes da Universidade Federal da Paraíba (UFPB) e características dos docentes. Notadamente, esta seção esclarece a estratégia adotada para atingir o objetivo proposto, no qual inicialmente, apresentam-se a abordagem empírica adotada e, posteriormente, expõem-se os dados utilizados na construção do painel quantílico.

1.3.1 Estratégia empírica

A preocupação central deste ensaio concentra-se na estimação do efeito distributivo do tamanho da classe no ensino universitário, a especificação do modelo segue a literatura existente (Machado e Vera-Hernandez, 2008; Bandiera, Larcinese e Rasul, 2010;), em especial ao adotado em De Giorgi et. al. (2012) e Harding e Larmarche (2014)¹⁰:

$$y_{ict} = \beta' x_{ict} + \eta_{it} + u_{ict} \quad (1)$$

onde y_{ict} é a nota média do aluno i , na classe c , período t ; x_{ict} é o vetor de variáveis exógenas que inclui o tamanho da turma, proporção de alunos do sexo feminino, proporção de alunos com renda alta e nota de entrada do aluno i . Por fim, o termo η_{it} que corresponde os fatores não observados associados ao aluno i ; e período t . Importante destacar que a equação (1) é a identificação adequada para o problema de pesquisa, pois se considera a existência da endogeneidade, mas admitindo-a que ela é controlada no tempo.

De Giorgi et. al (2012) e Harding e Lamarche (2014) consideram as variáveis: tamanho da turma; % aluno do sexo feminino e % de aluno com renda alta como endógeno. No entanto, para o caso do sistema público de ensino superior do Brasil, particularmente a UFPB, as variáveis são exógenas ao modelo, veja:

- a) Tamanho da turma – no sistema público de ensino do Brasil, os alunos não escolhem as disciplinas formando portfólios como nas universidades fora do país. O aluno ao ingressar na universidade se depara com uma oferta restrita de disciplinas que compõem seu currículo acadêmico e que normalmente é ofertada uma disciplina por semestre. Apesar de existir certa flexibilidade, no geral o tamanho da turma é exógeno.
- b) Proporção de alunos com renda alta – ao ingressar na UFPB, alunos ricos ou pobres se deparam com a mesma oferta de disciplinas. Por exemplo, não existem professores que montem cursos mais caros, logo, independente da classe social, os mesmo cursos são ofertados. Destarte, proporção de alunos de renda alta também é exógeno.
- c) Proporção de alunos do sexo feminino – é análoga a análise anterior, pois o fato do discente ser mulher não implica matrículas em disciplinas diferenciadas, logo, também é exógeno.

¹⁰ Harding e Lamarche (2014) consideram um modelo estrutural, uma vez que considera que existe um vetor de variáveis potencialmente endógenas.

Neste arcabouço, a equação (1) é a Função de Produção¹¹ que incorpora a heterogeneidade não observada, permitindo que variáveis latentes afetem o desempenho acadêmico e não contabilizá-las poderia causar viés nas estimações. Para atingir o objetivo proposto por este trabalho, esta seção apresenta o estimador de dois estágios abordado em Canay (2011).

I - Estimador de Regressão Quantílica com efeitos fixos – QR(FE)

O enfoque dado neste artigo permite que as covariadas afetem tanto a posição quanto a dispersão e o formato da distribuição dos desempenhos acadêmicos dos alunos. Dessa maneira, temos o impacto das covariáveis para cada ponto da distribuição. A operacionalização é dada pelo método de regressão quantílica. Por outro lado, os dados em painel permite incluir os efeitos fixos para controlar por fatores não observados. O método abordado em Canay (2011), potencialmente abarca tais especificações e o método aqui proposto.

Canay (2011) apresenta um estimador consistente em que utiliza uma simples transformação de dados para eliminar os efeitos fixos dos *quantis* condicionais, conforme o modelo:

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it} \quad E(\varepsilon_{it}/X_{it}, \alpha_i) = 0 \quad (2)$$

Onde $(y_{it}, x_{it}) \in \mathbb{R} \times \mathbb{R}^k$ são variáveis observáveis, $\theta(U_{it}) = \beta$ e $(U_{it}, \alpha_{it}) \in \mathbb{R} \times \mathbb{R}$ são não observados. $U_{it}/x_{it}, \alpha_{it} \approx U[0,1]$. A função $\tau \rightarrow x'\theta(\tau)$ é assumida como sendo estritamente crescente em $\tau \in (0,1)$ e o parâmetro de interesse é assumido para ser $\theta(\tau)$. A principal restrição é que as heterogeneidades individuais associadas a x_{it} deve ter um efeito local puro.

Em suma, o método versa na questão em que as condições (U_{it}, α_{it}) do parâmetro $\theta(\tau)$ podem ser identificadas e consistentemente calculadas a partir dos dados. Para estimar a equação (2), Canay (2011) propõe um estimador simples de dois passos, sendo $E_T(\cdot) \equiv T^{-1} \sum_{t=1}^T (\cdot)$ e $E_{nT}(\cdot) \equiv (nT)^{-1} \sum_{t=1}^T \sum_{i=1}^n (\cdot)$, como segue:

Passo 1. Deixe $\hat{\theta}_\mu$ para $\sqrt{nT}$ – estimador consistente de θ_μ . Estimam-se os efeitos fixos como:

$$\hat{\alpha}_i \equiv E_T[y_{it} - x'_{it}\hat{\theta}_\mu] \quad (3)$$

Passo 2. Sendo $\hat{y}_{it} \equiv y_{it} - \hat{\alpha}_i$ e definir o estimar do dois-estágios $\hat{\theta}(\tau)$ como:

$$\hat{\theta}(\tau) \equiv \operatorname{argmin}_{\theta \in (\cdot)} E_{nT}[\rho_r(\hat{y}_{it} - x'_{it}\theta)] \quad (4)$$

¹¹ A Função de Produção de Educação clássica, denominada de FPE, analisa a forma como diversos insumos do processo educacional podem afetar os resultados educacionais dos indivíduos. É dada por $Y_{it} = g(F_i^t, P_i^t, I_i, S_i^t)$, onde Y é um vetor educacional realizado pelo estudante; F é um vetor de características individuais e familiares; P é um vetor do corpo discente (influência dos pares); I é um vetor de dotações iniciais; e S é um vetor de insumos escolares relevantes (Hanushek, 1970, 1979, 2007).

Grosso modo, a ideia do método consiste em apurar o efeito fixo da variável dependente através de um estimador de média condicional (OLS), e depois aplicar uma regressão quantílica em dados em painel substituindo a variável dependente pelo valor deduzido do efeito individual.

1.3.2 Descrição dos dados

As estimativas da pesquisa foram possibilitadas pelos registros acadêmicos estudantis compilados pelo Núcleo de Tecnologia da Informação (NTI) e pela Comissão Permanente de Vestibular (COPERVE). A partir das duas fontes de dados foi possível obter informações do desempenho dos alunos ao longo do curso, bem como características socioeconômicas dos ingressantes nos cursos de graduação da Universidade Federal da Paraíba (UFPB), Campus I, durante o período de 2001 à 2010.

Do NTI foram extraídos variáveis sobre o desempenho acadêmico dos alunos ao longo do curso expressas pelas notas das disciplinas, bem como informações das respectivas disciplinas cursadas, o número de matriculados, reprovados, trancados, aprovados e o coeficiente de rendimento final (CRE)¹², entre outros. Além disso, a base ainda possibilita obter informações sobre os docentes, como carga horária e titulação.

Tabela 1.3 – Estatísticas Descritivas Geral

Variáveis	Nº Obs	Média	Desvio Padrão	Mínimo	Máximo
Desempenho Acadêmico	196.973	6.29	3.43	0	10
Tamanho da Turma	196.973	26.45	15.69	1	95
Proporção de alunos do sexo feminino	196.973	48.91	26.48	0	100
Proporção de alunos com renda alta	196.973	34.10	28.02	0	100
Nota de entrada (Vestibular)	196.973	532.60	61.68	390.2	752

Fonte: Elaboração Própria a partir dos dados do NTI (UFPB).

Por sua vez, os dados da COPERVE compreendem o questionário socioeconômico respondido pelo ingressante no momento do vestibular. Destarte, as informações concernem características do aluno como sexo, raça, local de nascimento, rede escolar ao qual o aluno estudou no ensino fundamental e médio, além do *background* familiar como escolaridade do pai, nível de ocupação, como também o resultado no teste de inscrição, ou seja, a nota final

¹² De acordo com a **Resolução UFPB nº 47/1999**, o Coeficiente de Rendimento Escolar (CRE) do aluno é obtido a partir da seguinte fórmula: $CRE = \frac{\sum M_i D_i}{C_h}$, onde M_i é a média final da *i-ésima* disciplina cursada ou dispensada; D_i é o número de horas-aula correspondente a *i-ésima* disciplina cursada ou dispensada; e, C_h é o número total de horas-aulas o que equivale ao número de créditos total correspondente às disciplinas cursadas ou dispensadas excluindo as disciplinas trancadas.

obtida no vestibular. As Tabelas 1.3 e 1.4 reportam-se a algumas estatísticas descritivas sobre as variáveis selecionadas da amostra:

A Tabela 1.3 apresenta a média, desvio padrão, valor mínimo e valor máximo das variáveis selecionadas neste estudo de um modo geral considerando um painel aluno-disciplina-período. No total são 196.973 observações referentes a 6.167 alunos com suas respectivas notas em 47.375 disciplinas ofertadas ao longo do tempo. A nota média dos alunos é em torno de 6,30 representativas de perfil discente no qual em média maior proporção de alunos são do sexo feminino, cerca de 49%, e apenas 34%, em média, possuem renda alta. Além disso, é possível notar que o tamanho médio das turmas é aproximadamente 26 alunos por classe. No entanto, se dividirmos os dados por centro acadêmico, observa-se que as médias das variáveis selecionadas divergem, denotando diferentes perfis de aluno para cada área de curso, conforme a Tabela 1.4.

Tabela 1.4 – Estatística Descritiva por Centro Acadêmico – Média

Variáveis/Centros Acadêmicos	CCEN	CI	CT	CCSA	CCHLA	CCTA	CE	CCS	CCJ	CCM
Desempenho Acadêmico	4.91	5.49	5.53	5.91	6.62	7.06	7.39	7.47	8.29	8.48
Tamanho da classe	34.96	35.65	36.18	42.22	38.56	28.66	38.31	33.01	38.84	35.93
Proporção do sexo feminino	34.72	15.35	34.5	43.02	64.05	54.01	88	64.08	43.03	43.85
Proporção de alunos com renda alta	23.21	60.52	43.41	33.57	26.31	34.38	7.96	27.8	78.87	70.48
Nota de entrada	506.81	608.45	549.2	513.48	509.93	519.97	463.78	555.25	647.53	697.9
Nº Observações	29665	3474	28736	34460	24802	14188	10764	31463	3918	4953

Fonte: Elaboração Própria a partir dos dados do NTI (UFPB).

Da Tabela 1.4 é possível apontar algumas peculiaridades dos cursos oferecidos pela instituição. Notem que os centros CCM e CJ, dos quais ofertam os cursos de medicina e direito, respectivamente, tem maior nota média de entrada como também maiores desempenhos acadêmicos médios, além do que a maior proporção de alunos com renda alta. Com relação a proporção de alunos do sexo feminino, o maior percentual médio encontra-se no Centro de Educação corroborados pelos cursos de pedagogia. No que tange o tamanho médio das turmas, averigua-se uma uniformidade na média, pois são cerca de 30 a 40 alunos por turma.

Outro fato interessante corresponde ao Centro de Tecnologia, onde estão os cursos de engenharia, note que apesar da nota média de inscrição ser alta, esta não é acompanhado pelo desempenho acadêmico dos alunos. Veja que a nota média é de 5,53, uma das mais baixas, perdendo apenas para a nota média do CCEN (onde ficam cursos como física, química, etc.) do CI (ciências da computação, matemática computacional). Parece que a entrada para esses

centros é difícil dada a nota média do vestibular ser alta, mas manter este desempenho ao longo do curso parece ser ainda mais difícil.

É imperioso destacar que o olhar breve por essas estatísticas descritivas não é possível identificar se há uma relação negativa entre o tamanho da turma e as notas das disciplinas. Isto se deve ao um conjunto de fatores associados aos indivíduos, ou docentes, ou a instituição ou mesmo ao tempo que podem influenciar positivamente ou negativamente nesta relação. Desta feita, na próxima seção aplica-se a metodologia adequada para obter estimativas confiáveis desta relação.

1.4 RESULTADOS E DISCUSSÃO

Nesta seção analisa-se o efeito do tamanho da classe no desempenho acadêmico. Para isso estima-se a equação (1) pelo método abordado por Canay (2011) no qual se considera a heterogeneidade não observada e os efeitos fixos. Posteriormente, restringe-se a amostra para aqueles indivíduos pertencentes aos centros acadêmicos a fim de captar peculiares efeitos no desempenho dos discentes. Finalmente, investiga-se a interação das covariáveis para testar seu impacto no desempenho acadêmico.

1.4.1 Linha base

Com o objetivo de investigar como o tamanho da turma influencia a *performance* dos alunos nas disciplinas cursadas, realizam-se regressões em um painel *aluno-disciplina-período* que incluem estudantes entre o período de 2001 à 2010, utilizando variações de controles para discentes e para docentes. Os controles individuais correspondem a características dos indivíduos tais como: *i)* sexo, *ii)* cor, *iii)* se havia exercido trabalho remunerado no momento da inscrição do vestibular, *iv)* se possuía computador em casa, *v)* rede escolar do ensino fundamental; e *vi)* rede escolar do ensino médio. Além disso, variáveis relacionadas ao *background* familiar como: *vii)* escolaridade do pai¹³, *viii)* se os pais trabalham, e *ix)* renda familiar. Quanto ao controle dos professores tem-se e a carga horária trabalhada no período.

Os resultados reportados na Tabela 1.5 correspondem a diferentes estimadores e especificações do modelo. Para fins de comparação, utilizou-se a mediana ($\tau = 0.5$) para obter as estimativas das colunas 1 à 3 que correspondem ao estimador padrão de regressão quantílica (QR) e, nas colunas 4 à 6 ao estimador robusto de Canay (2011) – Regressão Quantílica para estrutura de dados em painel. A variação entre as colunas para cada estimador correspondem ao uso de controles individuais e de professores. Em síntese, observam-se as estimativas foram estatisticamente significativas a pelo menos 1%, revelando *a priori* uma relação negativa entre o tamanho da classe e o resultado educacional dos indivíduos (Bandiera et. al., 2010; Machado Pinto e Vera Hernandez, 2010; De Giorgi, et. al, 2012).

Além disso, deduz-se que mudanças na proporção de mulheres afetam positivamente a variável resposta, ou seja, estudantes do sexo feminino tendem a melhorar o desempenho mediano da turma (Monks e Schmidt, 2010; Harding e Lamarche, 2015). No entanto, o mesmo não ocorre para a proporção de indivíduos com renda alta, note que os parâmetros foram estatisticamente insignificantes e parece que ter mais alunos com renda alta não afeta o desempenho acadêmico mediano (Harding e Lamarche, 2014). No que diz respeito a variável nota de entrada, ou seja, nota no exame de vestibular, verifica-se que há um efeito estatisticamente significativo e positivo, do qual alunos com testes de entrada mais altos tendem a melhorar o desempenho dos estudantes na mediana (Harding e Lamarche, 2014).

¹³ Optou-se pela escolaridade do pai, pois a base não continha informações sobre a escolaridade da mãe.

Tabela 1.5 – Resultados Empíricos para a mediana $\tau = 0.5$

	QR			QR(FE)		
	(1)	(2)	(3)	(4)	(5)	(6)
Efeitos de Interesse						
Tamanho da Classe	-0.0073*** (0.0005)	-0.00847*** (0.00046)	-0.0087*** (0.0005)	-0.0050*** (0.0004)	-0.0062*** (0.0003)	-0.0056*** (0.0004)
Proporção do sexo feminino	0.0195*** (0.0004)	0.02306*** (0.00030)	0.0195*** (0.0004)	0.0251*** (0.0003)	0.0313*** (0.0002)	0.0038*** (0.0258)
Proporção de renda alta	-0.0001 (0.0004)	-0.00077* (0.00032)	-0.0005 (0.0004)	0.0002 (0.0003)	-0.0005* (0.0002)	-0.0004 (0.0003)
Nota de entrada	0.0069*** (0.0002)	0.00616*** (0.00014)	0.0068*** (0.0002)	0.0119*** (0.0001)	0.0112*** (0.0001)	0.0121*** (0.0001)
Controles						
Características individuais	sim	não	sim	sim	não	sim
Características professores	não	sim	sim	não	sim	sim
Observações	145.692	138.185	100.976	145.692	138.185	100.976

Fonte: Elaboração Própria a partir dos dados do NTI (UFPB).

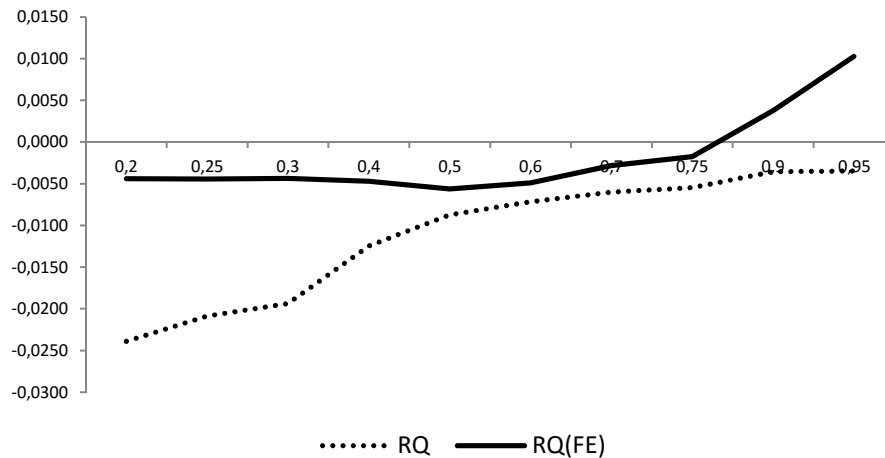
Nota 1: Erros padrão robustos entre parênteses. ***p-valor < 0,01. ** p-valor < 0,05. * p-valor < 0,10.

Nota 2: As características individuais são: i) sexo; ii) raça; iii) renda familiar; iv) escolaridade do pai; v) trabalho remunerado do pai e da mãe; vi) *dummy* se o aluno estudou em escola pública no ensino fundamental; vii) *dummy* se o aluno estudou em escola pública no ensino médio; viii) *dummy* se o aluno exercia algum trabalho remunerado no época do vestibular; ix) *dummy* se o aluno tinha computador em casa. As características dos professores são: i) titulação; ii) carga horária.

Em termos de magnitude, a coluna (3) revela que um aluno adicional na classe reduz o grau médio dos estudantes em cerca de 0,009. No entanto, para a estimativa RQ(FE) – coluna (6) é um pouco menor esta magnitude, cerca de -0,006. Apesar da diferença pequena é importante frisar o método QR não consegue captar características fixas não observáveis que determinam o desempenho acadêmico dos indivíduos. Harding e Lamarche (2015) chama atenção para a heterogeneidade latente existente no modelo aqui estimado, como por exemplo, a motivação e capacidade do aluno em absorver conhecimento em sala de aula.

É interessante notar que a partir das inferências apresentadas na Tabela 1.5, poder-se-ia concluir que políticas de redução de número de alunos por turma teria um efeito benéfico no desempenho acadêmico dos alunos da UFPB. Porém, se avançarmos um pouco mais na análise, indo além das estimativas condicionais a mediana com o intuito de analisar os resultados ao longo de toda a função densidade, verifica-se que isso não é verdade para todos da distribuição. O Gráfico 1.1 apresenta a evolução dos coeficientes da variável de interesse com um intervalo de confiança de 1% para todos os *quantis*, comparando a estimativa para regressão quantílica padrão.

Gráfico 1.1 – Estimativas do impacto do tamanho da classe no desempenho acadêmico dos indivíduos ao longo da distribuição



Fonte: Elaboração Própria a partir dos dados do NTI (UFPB).

Nota: Todas as estimativas foram estatisticamente significativas a 1%.

A linha pontilhada corresponde aos parâmetros do estimador de regressão quantílica padrão, enquanto que a linha contínua corresponde ao estimador de Canay (2011) utilizando como *proxy*¹⁴ para desempenho as notas das disciplinas cursadas. Note que com um intervalo de confiança de 99% associado, os parâmetros de estimador RQ tem: *i*) tamanho do efeito maior, em valor absoluto, nos *quantis* mais baixos, e *ii*) negativo em cada *quantil* (τ). Neste contexto, dois pontos merecem nota. Primeiro, quando aplica-se o estimador de Canay (2011) averigua-se que existe um *quórum* demasiado nos baixos *quantis* e isso se deve a omissão de efeitos fixos não observados que geram viés nas estimações.

Segundo, complementa a implicação do primeiro ponto. Sabe-se que o viés impede medir o correto impacto de políticas de redução de turmas. Contabilizadas a heterogeneidade não observada e os efeitos fixos, o estimador RQ(FE) avalia que redução do tamanho da turma teria um efeito positivo apenas para os baixos *quantis*, enquanto que este tipo de política prejudicaria alunos pertencentes aos *quantis* mais altos de performance.

Harding e Lamarche (2014) lançam luz com algumas explicações plausíveis sobre heterogeneidade desses efeitos, pois argumentam que classes menores prejudicariam alunos que tem boas notas nas disciplinas, uma vez que estes conseguem aprender mais com seus pares. Como contrapartida, classes menores poderiam favorecer aqueles alunos de baixo desempenho porque estes poderiam se beneficiar de maiores interações com seus professores, portanto, ajudando-os a melhorar o desempenho. Logo, os efeitos distributivos do desempenho dos alunos da UFPB variam de acordo com o tamanho da classe, e medir seu correto impacto é essencial para as políticas públicas.

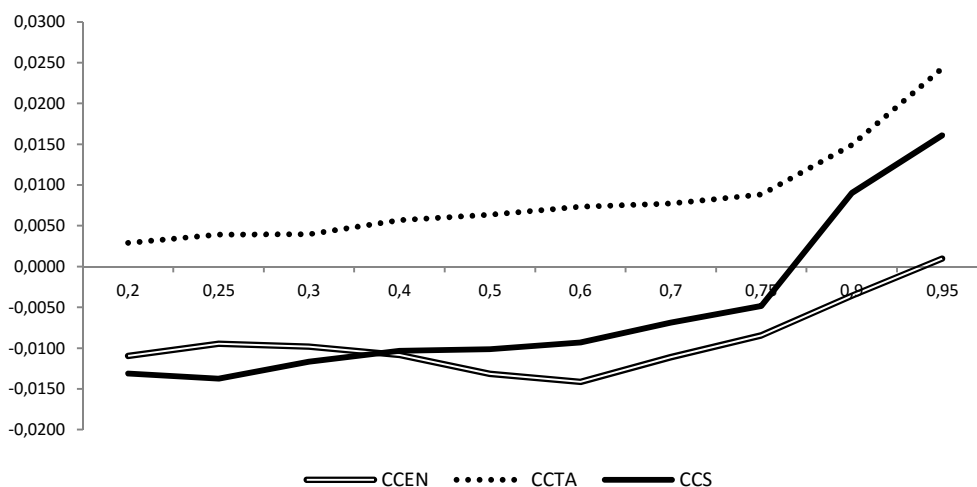
¹⁴ Os efeitos do tamanho da turma também foram testadas para CRE final como uma *proxy* alternativa de desempenho acadêmico, conforme apresentado no Apêndice B. Avaliando sob esta nova métrica, observa-se os mesmo resultados, concluindo que o método é consistente independente da *proxy*.

1.4.2 Política de Redução do Tamanho da turma por Centro Acadêmico

Utilizando a metodologia do Canay (2011), o mesmo exercício fora realizado para centros acadêmicos. Para averiguar o efeito do tamanho da turma selecionou-se o Centro de Ciência de Saúde (CCS), Centro de Ciência Exatas e da Natureza (CCEN) e Centro de Comunicação, Turismo e Artes (CCTA), dos quais apresentaram o melhor e o pior desempenho médio, e aquele com menor tamanho de turma, respectivamente, apresentados no Gráfico 2.

O gráfico 1.2 revela algumas evidências peculiares. Primeiro, apreende-se que o efeito heterogêneo ao longo da distribuição do tamanho da turma não é consensual entre os centros acadêmicos, notem que o CCTA (linha pontilhada do gráfico) apresenta estimativas estatisticamente significativas e positivas ao longo de todos os *quantis*. Há duas explicações plausíveis para este indício. Partindo do pressuposto de que alunos de alto desempenho tendem a se beneficiar com turmas maiores (Harding e Larmarche, 2014), é importante observar que o CCTA apresentou um os maiores desempenhos médios da amostra (conforme Tabela 4).

Gráfico 1.2 – Estimativas ao longo da distribuição por centro acadêmico – CCEN, CCTA e CCS



Fonte: Elaboração Própria a partir dos dados do NTI (UFPB).

Nota: Todas as estimativas foram estatisticamente significativas a pelo menos 1%.

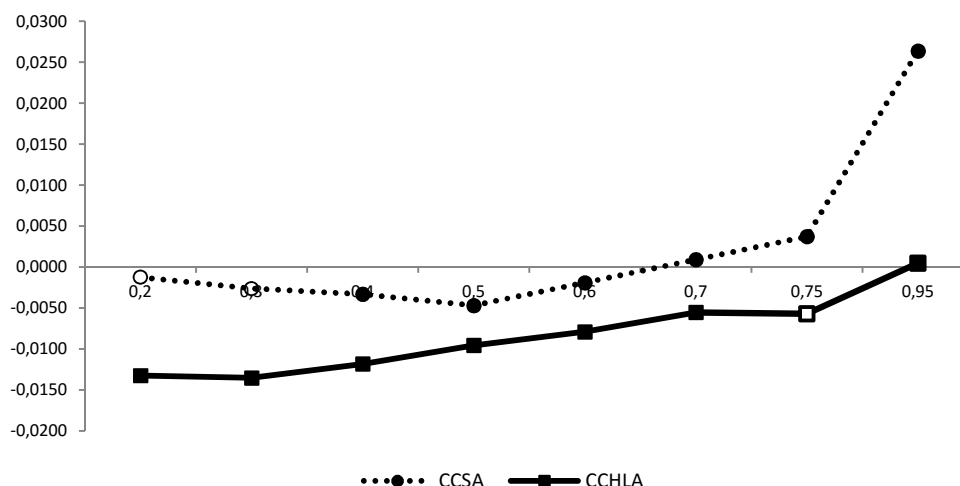
Além disso, sabe-se que esse benefício é devido às habilidades que os indivíduos possuem de aprenderem com seus pares. Sabe-se que a composição de pares é um importante determinante do comportamento individual e, conseqüentemente, do desempenho dos alunos (De Giorgi et. al., 2009), além da concorrência atuando como um forte dispositivo emocional (Harding e Lamarche, 2014). Ora, as interações entre pares são fundamentais para cursos como artes cênicas, música, jornalismo, turismo, cinema, entre outros, cursos estes pertencentes ao CCTA, então seria razoável pensar que quanto maiores as turmas, mais interações entre pares, reforçados por melhores desempenhos médios, trariam benefícios para

qualquer *quantil* da distribuição. Logo, política de redução de turma tenderia a prejudicar o desempenho dos estudantes deste centro.

Segundo ponto diz respeito às estimativas referentes ao CCEN e CCS. Conforme já apresentado na Tabela 1.4 deste artigo, o CCEN é o centro com menor desempenho médio e o CCS o maior. Temos dois casos distintos e mesmo resultado qualitativo, que seja, políticas de redução de turma tenderia a melhorar o desempenho dos alunos mais fracos e prejudicar alunos mais fortes. Para os dois centros há heterogeneidade de efeitos.

No Gráfico 1.3, selecionaram-se dois centros da área de humanas: CCHLA e CCSA, do qual o curso de Ciências Econômicas faz parte deste último. Análogo as análises anteriores, reforça-se que o impacto do tamanho da classe diminui negativamente no medida que cresce o desempenho acadêmico dos alunos. Em outras palavras, é possível perceber que o efeito é negativo nos baixos *quantis*, com tendência a diminuir na medida em que avançamos nos *quantis*. Outrossim, a eficácia de uma política de redução teria diferentes impactos para os dois centros, já que para o CCSA o tamanho da turma tem efeito positivo em *percentis* mais baixo (a partir do *quantil* 70) do que o CCHA (a partir do quantil 90).

Gráfico 1.3 – Estimativas ao longo da distribuição por centro acadêmico – CCSA, CCHLA



Fonte: Elaboração Própria a partir dos resultados das estimações.

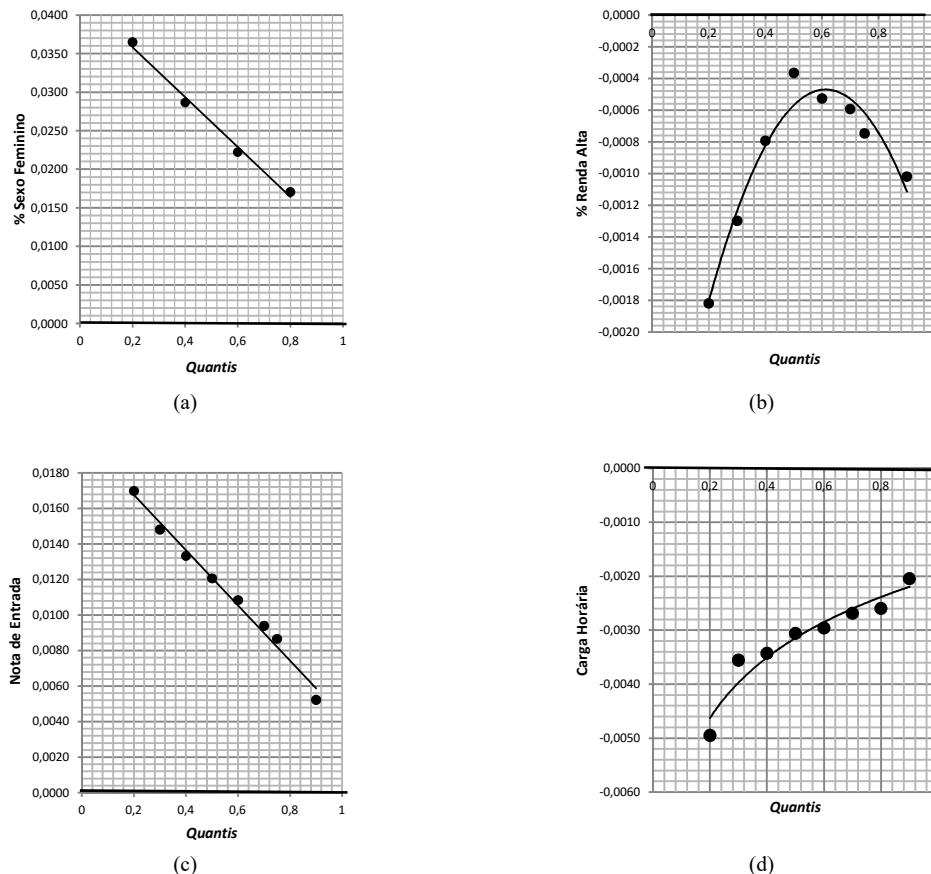
Nota: Para representar a significância dos parâmetros utilizou-se o marcador fechado e aberto. Marcador fechado correspondem as estimativas estatisticamente significativas pelo menos 1%, já o marcador aberto são parâmetros insignificantes estatisticamente.

É importante frisar que a constatação de efeitos heterogêneos é demasiadamente importante para os chefes de departamento, uma vez que existe um *trade-off* entre bons e maus alunos no objetivo de melhorar o desempenho médio do curso. A literatura precedente a Harding e Lamarche (2014) afirma que uma política de redução de turmas seria benéfica para todos da distribuição, e aqui se observa que não é bem assim. Aplicar uma política de redução de turmas pode comprometer resultados como um todo e particularmente danoso para determinados perfis de turmas.

1.4.2 Efeitos das Covariáveis

Além da variável de interesse tamanho da turma, é possível investigar alguns resultados quantílicos das covariáveis: proporção de alunos do sexo feminino; proporção de alunos com renda familiar alta e nota de entrada e carga horária. Os resultados obtidos são sumarizados no Gráfico 1.4.

Gráfico 1.4 – Efeitos quantílicos das covariáveis sobre o desempenho acadêmico.



Fonte: Elaboração Própria a partir das estimativas.

Nota: Todas as estimativas foram estatisticamente significativas.

Em cada gráfico, os pontos pretos ilustram o valor do coeficiente da covariável correspondente para cada quantil da distribuição de desempenho educacional. A linha contínua em torno dos pontos representa a tendência dos efeitos associados a cada covariável. Para facilitar a visualização e interpretação, a linha horizontal contínua em negrito divide os gráficos no ponto zero. O gráfico (a) corresponde aos resultados para a proporção do sexo feminino na sala de aula. Observa-se um efeito positivo para todos os *quantis*, que qualitativamente é análogo ao apresentado para a mediana na Tabela 5 e corrobora com os resultados de Monks e Schmidt (2010). No entanto, à medida que o desempenho aumenta menor é o efeito de mulheres adicionais na turma conforme apontado em Harding e Lamarche (2015).

O gráfico (b) mostra os resultados para a proporção de alunos com renda alta. Na Tabela 5, ao observar apenas os valores da mediana, concluiu-se *a priori* que adicionar alunos de renda alta nas turmas não afetaria o desempenho acadêmico. Todavia, este resultado não é ratificado ao longo da distribuição condicional. Observa-se que num intervalo de confiança de 90% associado, adição de alunos com renda alta tem efeito negativo ao longo da distribuição, sobretudo na cauda inferior. Por sua vez, o gráfico (c) traz as estimativas da nota de entrada dos indivíduos, ou seja, as notas nos exames do vestibular. Há um efeito estatisticamente significativo, positivo e decrescente ao longo da dos *quantis*, reforçando que estudantes que estavam com maiores notas no exame de entrada tem efeito sobre aqueles pertencentes aos baixos *quantis*.

Por fim, o gráfico (d) corresponde a variável de controle utilizada para captar as características de professores. Utilizou-se a carga horária exercida por semestre. Estatisticamente significativos, os resultados sugerem que a carga horária tem efeito negativo sobre o desempenho acadêmico dos indivíduos independente de sua posição na distribuição condicional. Veja-se que este efeito é ainda mais intenso na calda inferior. Em consonância com a análise do tamanho da turma, do qual parte-se do pressuposto que classes menores beneficiariam alunos mais fracos, isto porque teriam uma atenção adicional do professor (Haridng e Lamarche, 2014). Logo, quanto maior a carga horária do docente, menos tempo ele tem para se dedicar a produção de material e disponibilidade de atendimento a esses alunos.

1.5 CONSIDERAÇÕES FINAIS

Em relação à literatura existente, este artigo acrescenta a utilização de regressão quantílica para uma estrutura de dados em painel contornando a heterogeneidade não observada em ambientes de efeitos fixos para estimar a relação causal do tamanho da turma nos desempenhos acadêmicos do ensino universitário brasileiro, dado a inexistência de trabalhos com essa temática para o Brasil. Para além disso, realizam-se variações da amostra para analisar o impacto das políticas de redução de turmas e aferir-se os efeitos dos covariáveis estudadas.

Para tanto, optou-se pelo método abordado em Canay (2011) aplicado para um banco de dados único referente aos registros administrativos da UFPB compilados pelo NTI e COPERVE. Muito embora não exista um consenso em relação ao efeito do tamanho da turma, o pressuposto básico requerido é de que essa relação é negativa e que existem variáveis latentes como motivação e habilidade dos indivíduos que interferem nessa relação. Dessa forma, o método contorna estes problemas e contribui fornecendo estimativas do impacto do tamanho da classe no desempenho acadêmico, medida pelas notas obtidas ao final das disciplinas.

Os resultados gerais sugerem que uma política de redução de turma tem efeito positivo para aqueles indivíduos pertencentes aos baixos *quantis*, ou seja, aqueles indivíduos com menor desempenho acadêmico. Por sua vez, aqueles indivíduos com maiores desempenhos, uma política eficaz seria aumentar o número de alunos na classe. A explicação para isso é dada em Harding e Lamarche (2014) ao inferir que alunos com notas mais altas se beneficiam das interações com seus pares por possuírem maiores habilidade de absorção de conhecimento. Portanto, a conclusão é de que há efeitos heterogêneos ao longo da distribuição e aplicação de políticas de redução de turma pode prejudicar bons alunos em detrimento de beneficiar alunos mais fracos.

Além disso, as estimativas por centro acadêmico indicam que há diferentes efeitos entre os próprios centros de ensino. Esta constatação é plausível, uma vez que dentro da própria instituição encontram-se diferentes perfis de alunos e professores em razão das diversas áreas. Com isso, ressalta-se a existência de um *trade-off* no qual os chefes de departamento acadêmico se deparam, uma vez que escolher reduzir turmas, além de ser uma proposta cara, pois via de regra implica em mais insumos escolares para novas turmas, isso também pode significar comprometer desempenho acadêmico individual e consequentemente geral do curso. Em consonância com os demais resultados, as covariáveis apresentaram-se como o esperado, deduz-se que *i)* estudantes adicionais do sexo feminino tendem a melhorar da turma em geral; *ii)* adição de alunos com renda alta tem efeito negativo ao longo da distribuição, sobretudo na cauda inferior e, *iii)* efeito positivo para nota de entrada confirmando que alunos de bom desempenho no vestibular tem efeito assertivo para aqueles ainda em realização acadêmica.

Finalmente, os resultados aqui expostos podem servir de respaldo para os administradores universitários avaliarem melhor os efeitos do tamanho da classe no desempenho acadêmico dos estudantes da UFPB. Além do mais, em que pese a literatura limitada e a inexistência de trabalhos para o Brasil, é importante esclarecer que o trabalho também abre uma agenda de pesquisa relevante. A falta de consenso sobre essa variável já dá indícios da importância do tema.

1.6 REFERÊNCIAS

Alam, Mahmudul. Development of primary education in Bangladesh: the ways ahead. **The Bangladesh Development Studies**, p. 39-68, 2000.

Asadullah, M. Niaz. The effect of class size on student achievement: evidence from Bangladesh. **Applied Economics Letters**, v. 12, n. 4, p. 217-221, 2005.

Angrist, JOSHUA D .; LAVY, VICTOR. USANDO MAIMONIDES' RULE para estimar o efeito da classe SIZE ON desempenho escolar. **O Quarterly Journal of Economics**, p. 533-575, 1999.

Bandiera, Oriana; Larcinese, Valentino; Rasul, Imran. Heterogeneous class size effects: New evidence from a panel of university students*. **The Economic Journal**, v. 120, n. 549, p. 1365-1398, 2010.

Banerjee, Abhijit V. et al. Remedying Education: Evidence from Two Randomized Experiments in India. **The Quarterly Journal of Economics**, v. 122, n. 3, p. 1235-1264, 2007.

Bedard, Kelly; Kuhn, Peter. Where class size really matters: Class size and student ratings of instructor effectiveness. **Economics of Education Review**, v. 27, n. 3, p. 253-265, 2008.

Camargo, Juliana; Porto Júnior, Sabino. O efeito do tamanho da turma sobre o desempenho escolar: uma avaliação do impacto da "enturmação" no ensino fundamental do Rio Grande do Sul. Texto para discussão Nº 02, 2014.

Canay, Ivan A. A simple approach to quantile regression for panel data. **The Econometrics Journal**, v. 14, n. 3, p. 368-386, 2011.

De Giorgi, Giacomo; Pellizzari, Michele; Woolston, William Gui. Class size and class heterogeneity. **Journal of the European Economic Association**, v. 10, n. 4, p. 795-830, 2012.

De Paola, Maria; Ponzio, Michela; Scoppa, Vincenzo. Class size effects on student achievement: heterogeneity across abilities and fields. **Education Economics**, v. 21, n. 2, p. 135-153, 2013.

Duflo, Esther; Dupas, Pascaline; Kremer, Michael. **Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya**. National Bureau of Economic Research, Inc, 2008.

Harding, Matthew; Lamarche, Carlos. Estimating and testing a quantile regression model with interactive effects. **Journal of Econometrics**, v. 178, p. 101-113, 2014.

Hoxby, Caroline. **Peer effects in the classroom: Learning from gender and race variation**. National Bureau of Economic Research, 2000.

Krueger, Alan B. Experimental Estimates of Education Production Functions. **Quarterly Journal of Economics**, p. 497-532, 1999.

Machado, Matilde Pinto; Vera-Hernandez, Marcos. **Does class-size affect the academic performance of first year college students**. mimeo, University College London, 2008

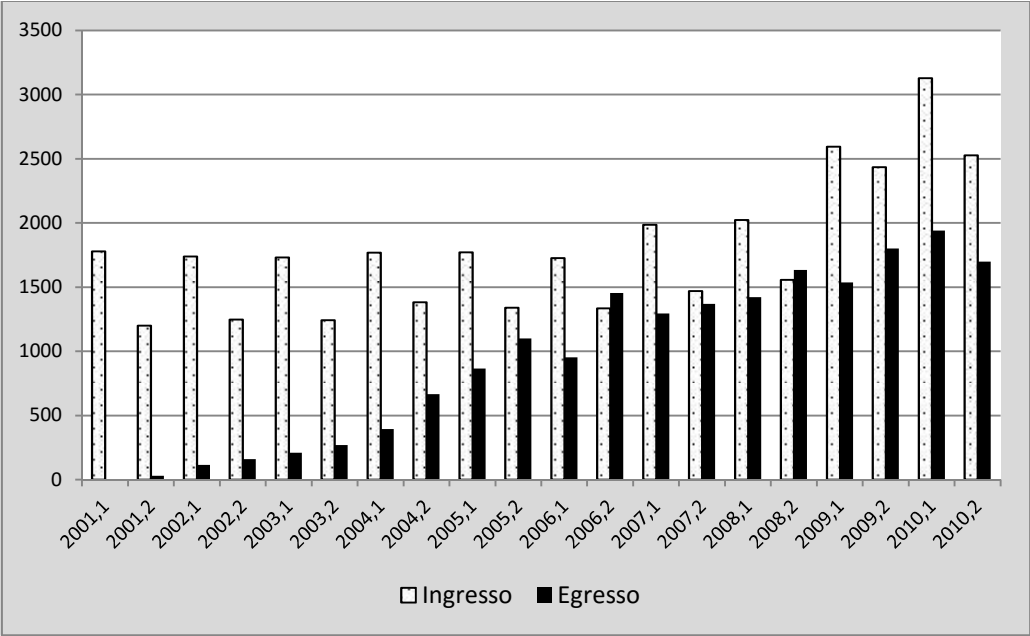
Menezes-Filho, Naércio Aquino. **Os determinantes do desempenho escolar do Brasil**. IFB, 2007.

Monks, James; Schmidt, Robert M. The impact of class size on outcomes in higher education. **The BE Journal of Economic Analysis & Policy**, v. 11, n. 1, 2011.

Oliveira, Jaqueline Maria de. Custo-efetividade de políticas de redução do tamanho da classe e ampliação da jornada escolar: uma aplicação de estimadores de *matching*. **Rio de Janeiro: Banco Nacional de Desenvolvimento Econômico e Social**, 2010.

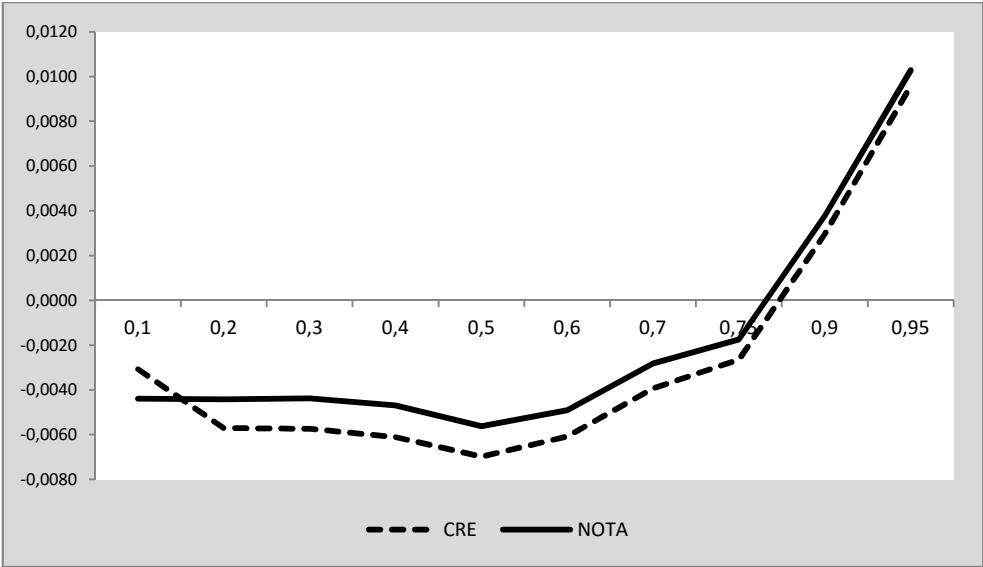
Uquiola, Miguel. Identifying class size effects in developing countries: Evidence from rural Bolivia. **Review of Economics and statistics**, v. 88, n. 1, p. 171-177, 2006.

APÊNDICE A1 – NÚMERO DE INGRESSOS E EGRESSOS – UFPB, 2013.



Fonte: Elaboração Própria a partir dos dados do NTI (UFPB).

APÊNDICE A2 – COMPARAÇÃO DE RESULTADOS DO EFEITO DA TURMA COM DUAS MÉTRICAS



Fonte: Elaboração Própria a partir dos resultados das estimações.

ANEXO 1 – LISTA DOS CENTROS ACADÊMICOS E OS CURSOS DE GRADUAÇÃO

1 - CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA (CCEN) 1.1 CIÊNCIAS BIOLÓGICAS 1.2 ESTATÍSTICA 1.3 FÍSICA 1.4 GEOGRAFIA 1.5 MATEMÁTICA 1.6 QUÍMICA 2 - CENTRO DE CIÊNCIAS HUMANAS, LETRAS E ARTES (CCHLA) 2.1 HISTÓRIA MSC 2.2 LETRAS CLÁSSICAS 2.3 CIÊNCIAS SOCIAIS 2.4 LETRAS LIBRAS 2.5 FILOSOFIA 2.6 HISTÓRIA 2.7 LETRAS 2.8 LÍNGUAS ESTRANGEIRAS APLICADAS ÀS NEGOCIAÇÕES INTERNACIONAIS 2.9 LETRAS VIRTUAL 2.10 MÍDIAS DIGITAIS 2.11 PSICOLOGIA 2.12 SERVIÇO SOCIAL 2.13 TRADUÇÃO 3 - CENTRO DE CIÊNCIAS MÉDICAS (CCM) 3.1 MEDICINA 4 - CENTRO DE EDUCAÇÃO (CE) 4.1 PEDAGOGIA 4.2 PEDAGOGIA - MODALIDADE À DISTÂNCIA 4.3 PEDAGOGIA - EDUCAÇÃO DO CAMPO 4.4 PSICOPEDAGOGIA 4.5 CIÊNCIAS DAS RELIGIÕES 4.6 CIÊNCIAS NATURAIS - MODALIDADE À DISTÂNCIA 5 – CENTRO DE CIÊNCIAS SOCIAIS APLICADAS (CCSSA) 5.1 ADMINISTRAÇÃO 5.2 CIÊNCIAS ECONÔMICAS 5.3 CIÊNCIAS CONTÁBEIS 5.4 ARQUIVOLOGIA 5.5 BIBLIOTECONOMIA 5.6 RELAÇÕES INTERNACIONAIS 6 - CENTRO DE TECNOLOGIA (CT) 6.1 ARQUITETURA E URBANISMO 6.2 ENGENHARIA AMBIENTAL 6.3 ENGENHARIA CIVIL 6.4 ENGENHARIA MECÂNICA 6.5 ENGENHARIA DE ALIMENTOS 6.6 ENGENHARIA DE MATERIAIS 6.7 ENGENHARIA DE PRODUÇÃO 6.8 ENGENHARIA DE MECÂNICA 6.9 ENGENHARIA QUÍMICA 6.10 QUÍMICA INDUSTRIAL 7 - CENTRO DE CIÊNCIAS DA SAÚDE (CCS) 7.1 CIÊNCIAS FARMACÊUTICAS 7.2 ENFERMAGEM 7.3 FISIOTERAPIA 7.4 FONODIOLOGIA 7.5 EDUCAÇÃO FÍSICA 7.6 NUTRIÇÃO 7.7 ODONTOLOGIA 7.8 TERAPIA OCUPACIONAL	8 - CENTRO DE CIÊNCIAS JURÍDICAS (CCJ) 8.1 DIREITO 9 - CENTRO DE BIOTECNOLOGIA (CBIOTEC) 9.1 BIOTECNOLOGIA 10 - CENTRO DE COMUNICAÇÃO, TURISMO E ARTES (CCTA) 10.1 ARTES VISUAIS 10.2 TEATRO 10.3 JORNALISMO 10.4 RELAÇÕES PÚBLICAS 10.5 RADIALISMO 10.6 CINEMA 11 - CENTRO DE ENERGIAS ALTERNATIVAS E RENOVÁVEIS (CEAR) 11.1 ENGENHARIA ELÉTRICA 11.2 ENERGIAS RENOVÁVEIS 12 - CENTRO DE INFORMÁTICA (CI) 12.1 CIÊNCIAS DA COMPUTAÇÃO 12.2 ENGENHARIA DA COMPUTAÇÃO 12.3 MATEMÁTICA COMPUTACIONAL 13 - CENTRO DE TECNOLOGIA E DESENVOLVIMENTO REGIONAL (CTDR) 13.1 TECNOLOGIA DE ALIMENTOS 13.2 TECNOLOGIA SUCRAALCOOLEITA
--	--

Fonte: Site da UFPB disponível em: <http://www.ufpb.br/content/centros-de-ensino>.

CAPÍTULO 2: Escassez de água e os Resultados Educacionais dos Universitários do Semiárido Brasileiro

Resumo

O presente artigo tem por objetivo investigar os efeitos das mudanças climáticas, corroborada pelos longos períodos de secas no nordeste, no desempenho acadêmico dos jovens universitários nascidos no semiárido brasileiro. Utiliza-se a função de tecnologia de habilidades abarcada no modelo teórico de Cunha e Heckman (2007), onde o estoque de capacidades é uma função das características dos pais, dos investimentos realizados pelos pais, das características genéticas e dos fatores ambientais aqui expressos na variável log-desvio da precipitação dos 12 meses anteriores ao nascimento do indivíduo. Desta forma, aplica-se a metodologia de regressão quantílica com efeitos fixos abordado em Canay (2011) em um painel *aluno-disciplina-ano*. Os resultados sugerem que quanto maior o nível de chuvas durante o período gestacional dos indivíduos nascidos no semiárido maior é seu impacto no desempenho acadêmico dos mesmos. Não obstante, averiguou-se que esse efeito é persistente e positivo ao longo da distribuição condicional de desempenho, principalmente nas extremidades das caudas.

Palavras-chaves: Precipitação; Estoque de Capacidades; Regressão Quantílica com efeitos fixos; Desempenho Acadêmico

Abstract

This article aims to investigate the effects of climate change, supported by long periods of drought in the northeast, in the academic performance of students in the Brazilian semiarid young. We use the skills of technology function encompassed the theoretical model of Cunha and Heckman (2007), where the stock capacity is a function of the characteristics of the parents, the investments made by the parents, genetic characteristics and environmental factors expressed here in variable log deviation of rainfall from 12 months prior to the birth of the individual. Thus, the quantile regression method with fixed effects discussed in Canay applies (2011) a panel-year student-subject. The results suggest that the higher the level of rainfall during pregnancy of individuals born in semiarid greater is its impact on the academic performance of the same. Nevertheless, it was established that this effect is persistent and positive throughout the conditional distribution performance, especially at the ends of the tails.

Keywords: precipitation; Capabilities of stock; Quantile regression with fixed effects; Academic performance

2.1 INTRODUÇÃO

A formação das habilidades cognitivas e não cognitivas são demasiadamente importantes nas explicações teóricas dos resultados educacionais e deliberações políticas sobre educação. Nesse contexto, Weiss (2010) avaliou os efeitos das habilidades cognitivas e não cognitivas nos ganhos dos indivíduos de diferentes sistemas escolares e concluiu que há um efeito positivo e especialmente maior na cauda superior da distribuição condicional de rendimentos. Buttner, Thiel e Thopson (2011) investigam o papel de uma mudança na intensidade de aprendizagem no ensino médio e constataam efeito positivo na formação de habilidades não cognitivas. Em suma, quando os indivíduos são submetidos a experimentos controlados, há uma persistência nos diferenciais de habilidade cognitiva e não cognitiva ao longo da vida e com diferentes intensidades para cada fase.

Recentes pesquisas tem se debruçado em investigar as implicações de restrições nos primeiros anos de vida sobre a formação de habilidade cognitiva e não cognitiva (Heckman, Stixrud e Uuzua, 2006; Cunha e Heckman, 2007). Com efeito, esta formação pode ser afetada por múltiplos e complexos fatores (Heckman, 2007), e embora possa ser afetada negativamente por herança epigenética, a baixa cognição pode ocorrer, também, pela ausência de condições favoráveis durante a gravidez e na infância (Currie, 2008).

Neste rol de discussões, Heckman, Stixrud e Urzua (2006) investigaram o efeito das condições iniciais sobre os determinantes do sucesso escolar, o crime, o lucro e, em seguida, são integrados em uma única abordagem teórica em Cunha e Heckman (2007). O autor reúne contribuições da literatura psicológica, da neurociência e da economia, particularmente no desenvolvimento de um quadro teórico agrupando habilidades cognitivas e não cognitivas no conceito de *capacidades*. A ideia é de que o indivíduo recebe influências genéticas, ambientais (sobretudo em experiência intrauterinas), investimentos ao longo da vida e herdabilidade parental, que compõem a formação de suas *capacidades* quando adulto.

Estas conexões permitem inferir possíveis relações entre as condições climáticas extremas e o desenvolvimento humano. Por exemplo, restrições pluviométricas propiciam ambientes iniciais desfavoráveis, uma vez que há efeito negativo em processos produtivos, na saúde e na qualidade de vida da região (Cirilo, 2008). A escassez de água pode reduzir a produção agrícola e a ingestão suficiente de nutrientes afetando os resultados de saúde, e a falta de água potável pode gerar maior incidência de doenças infecciosas comprometendo a qualidade de vida da região (Rocha e Soares, 2015). Não obstante, estas consequências são ainda mais danosas quando incidem sobre períodos intrauterinos e nos primeiros anos de infância por comprometerem a formação das habilidades cognitivas e não cognitivas dos indivíduos (Backer, 1998; Gluckman e Handson, 2005; Cunha e Heckman, 2007; Heckman, 2007).

Neste arcabouço teórico, o objetivo desta pesquisa é analisar a relação causal entre restrições pluviométricas no período gestacional sobre as habilidades cognitivas dos indivíduos quando adultos. Especificamente, procura-se observar se as habilidades dos indivíduos que nasceram no semiárido brasileiro foram afetadas pelas secas, discutindo as evidências encontradas na literatura sobre formação de habilidades cognitivas no contexto

brasileiro. Para isso, lança-se mão de uma painel *aluno-disciplina-ano* que acompanha os indivíduos que ingressaram na Universidade Federal da Paraíba no período de 2001 à 2010, possibilitados por um registro único de informações compiladas pelo Núcleo de Tecnologia da Informação (NTI) – UFPB, dos quais pôde-se obter um conjunto de variáveis relacionadas ao ambiente inicial, condições genéticas, atributos parentais, investimentos e desempenho nas disciplinas cursadas ao longo do curso.

Nada obstante, para a consecução deste objetivo é importante frisar que se utilizou como *proxy* das habilidades cognitivas quando adulto as notas das disciplinas, por permitir a construção de um painel *aluno-disciplina-ano*. Isto se deve a ausência de um banco de dados longitudinal no Brasil, ou seja, não há um banco que acompanhe temporalmente o indivíduo no momento da gravidez até adulto. Logo, a partir da data de nascimento foi possível conectar os dados de clima ao nascer e as informações socioeconômicas dos indivíduos e seus respectivos desempenhos. Além disso, aplicou-se o método abordado em Canay (2011) de Regressão Quantílica com efeitos fixos.

Além desta introdução, este trabalho está organizado da seguinte forma. Seção 2.2 apresenta-se uma breve revisão da literatura sobre a formação de habilidades dos agentes. A seção 2.3 expõe-se o modelo teórico e empírico, bem como o banco de dados. A seção 2.4 os resultados e a seção 2.5 as considerações finais. Finalmente, na seção 2.6 apresenta as referências bibliográficas.

2.2 FORMAÇÃO DE CAPACIDADES

“Smart investments start by addressing a major root cause of inequality - families and disadvantaged early childhoods.”

(James Heckman)

Nesta seção, objetiva-se apresentar um pouco da literatura sobre a formação de habilidades sob a ótica de quatro literaturas: psicologia, neurociência cognitiva, economia e epidemiologia. Há dois pontos centrais nesta discussão. O primeiro permeia o desenvolvimento teórico da formação da cognição dos indivíduos. A multidisciplinariedade teórica nos permite aproximar de um modelo mais realístico. O segundo é revelado pela frase acima de *James Heckman* de que uma forma de dirimir desigualdades existentes em famílias menos favorecidas deve começar nos primeiros anos de vida, isto é o mais inteligente.

2.2.1 Formação de Habilidades: Psicologia, Economia, Neurociência e Saúde

O sucesso socioeconômico dos indivíduos, comumente utilizado por um grande escopo de pesquisadores, é em larga medida determinado por características individuais. Estas características determinam o sucesso em diferentes dimensões de bem-estar e vem despertando grande interesse em investigar em que medida diferenças observadas de comportamento tem origem no desenvolvimento de habilidades. O grande desafio, contudo, é identificar a formação destas características e obter uma descrição do comportamento destas habilidades.

Na literatura da psicologia, Piaget (1982) destaca dois elementos básicos ao desenvolvimento cognitivo dos indivíduos: *i)* fatores invariantes; e *ii)* fatores variantes. O primeiro, conhecido na literatura como gênese, diz que ao nascer, o indivíduo recebe como herança uma série de estruturas biológicas (sensoriais e neurológicas) que permanecem constantes ao longo da vida. Por outro lado, o segundo fator expressa que o desenvolvimento cognitivo dos indivíduos se dá a partir da interação entre o homem e o meio em que ele vive.

Além disso, inspirados pelas possibilidades de medidas de inteligência consolidadas pela escala Binet-Simon¹⁵, Stanford-Binet¹⁶ (o teste de QI) e Spearman¹⁷ (o *fator g*), a literatura psicológica contribui neste rol de discussões ao postular que: *i)* há um processo cumulativo de conhecimento que varia com a idade dos indivíduos; *ii)* o conhecimento atual é insumo para o aprendizado posterior; e *iii)* a aprendizagem são captadas de forma distintas em diferentes fases da vida, no qual pode haver períodos em que este aprendizado pode ser facilitado.

¹⁵ Alfred Binet e Théodore Simon desenvolveram o primeiro teste moderno de inteligência, chamada de escala Binet-Simon (1905).

¹⁶ A escala Binet-Simon foi aprimorada pelo Professor Lewis Terman, passado a formar o mais popular teste de Quociente de Inteligência - QI (1916).

¹⁷ Charles Spearman desenvolveu um método estatístico, traduzido na teoria do fator *g*, em que provou que havia nos indivíduos um atributo capaz de captar a estrutura da inteligência.

A Neurociência vem a contribuir com a literatura a partir dos resultados do refutado debate conhecido como “*Nature versus Nurture*”. A partir dos estudos do cérebro e sua relação com a mente, à consciência e a cognição, distinguem que o ciclo da vida possui períodos críticos. Nestes períodos, quanto mais expostos a novas informações e atividades, mais circuitos neurais são criados e ampliados, e dependendo do ambiente, das circunstâncias e das experiências vividas, maiores serão suas diferenças comportamentais (Fiori, 2008). Assim, a partir da literatura epigenética, a literatura da neurociência cognitiva vem mostrar que medidas de habilidades são susceptíveis tanto por influências genéticas, mas também ambientais, inclusive em experiências intrauterinas.

Conquanto o desenvolvimento cognitivo dos indivíduos fosse amplamente difundido na literatura psicológica e a neurociência cognitiva, sua forte correlação com os ganhos de produtividade gerados pelo “fator humano” na produção despertou interesse entre os economistas protuberantes a formular a Teoria do Capital Humano (ver Mincer, 1958, 1974; Schultz, 1961, 1975; Becker, 1964, entre outros). O artigo de Schultz (1961) intitulado “*Investment in Human Capital*” foi uma das principais contribuições para a disseminação da teoria do capital humano. Em particular, para as pesquisas voltadas a investigar as relações de crescimento e desenvolvimento das nações.

Dentro da literatura econômica, a concepção do capital humano passou a ser documentado tanto do ponto de vista macroeconômico quanto microeconômico. O primeiro potencializando a produção gerando incremento de riqueza nas nações (Ver Romer, (1986, 1990); Lucas, 1988, Rebelo (1991), Mankiw, Romer e Weill (1992), Barro e Lee (2013). Por sua vez, na segunda ótica investigando os determinantes do capital humano para promover a própria existência econômica dos indivíduos (ver Becker e Tomes (1979), Hanushek (1970, 1979, 2007); Barros et. al., (2001); Checchi e Dardonani (2002); Ferreira e Veloso (2003)). No geral, a conclusão é de que melhorar a qualidade da força de trabalho é um elemento essencial para o objetivo de crescimento econômico.

Dentro deste escopo de pesquisa em economia, o tema transborda-se em várias outras áreas de estudo, sobretudo aquelas sob o enfoque de caracterizar o comportamento decisório dos agentes econômicos. Em suma, a formação e determinantes do comportamento dos atributos individuais, ou melhor, o desenvolvimento cognitivo dos indivíduos produz heterogeneidade entre os mesmos, aclarando os diferenciais de acumulação de capital humano (Heckman, 2007; Carneiro e Heckman, 2003) e produzindo importantes implicações em termos de políticas públicas.

Vale ressaltar que muito embora a literatura econômica tenha evoluído nas investigações do desenvolvimento cognitivo dos indivíduos, por exemplo, as pesquisas sobre desigualdade de oportunidades¹⁸, ainda é possível identificar duas grandes lacunas. Primeiro, ignora-se todo o desenvolvimento teórico da literatura psicológica em postular que existem

¹⁸ Baseado na Teoria da Justiça de Rawls (1971), Roemer (1993) postula que os resultados econômicos individuais são consequência de dois grupos: i) variáveis de circunstância (não responsabilidade) – as comumente encontradas na literatura de capital humano, ou seja, sexo, raça, *background familiar*, entre outras; ii) variáveis de esforço (responsabilidade) – cujo os indivíduos são responsáveis, por exemplo, decisão de migrar, anos de estudos, horas trabalhadas, etc.

vários períodos na infância. Na verdade, o pressuposto implícito é de que os *inputs* na produção de habilidades dos indivíduos são substitutos perfeitos (Cunha e Heckman, 2007), pois se considera que a tecnologia de produção de habilidades é idêntica e independente da idade dos indivíduos conferindo a formação de capital humano característica estática (Becker e Tomes (1986); Lam (1999); Barros et. al (2001)).

A segunda lacuna é decorrente da primeira. Investimentos tardios em habilidades podem não ser suficientes para remediar a carência de investimentos precoces, uma vez que existe um *trade-off* de equidade-eficiência para investimentos mais tarde e não para investimentos em estágios iniciais da vida (Cunha e Heckman, 2007). Além disso, alguns estudos tendem a negligenciar que as habilidades são de múltiplas naturezas, por exemplo, habilidade não cognitiva (como perseverança, aversão ao risco, motivação autoestima) e também tem efeitos nos resultados socioeconômico dos indivíduos (Heckman, 1995). Tais lacunas têm fortes implicações nas políticas públicas e sua existência na literatura econômica pode gerar resultados desastrosos no caminho da igualdade de oportunidades.

Portanto, na busca resumir as principais evidências e conclusões relacionadas às literaturas da psicologia, neurociências e economia, Cunha e Heckman (2007) desenvolvem um quadro teórico que parte do pressuposto que as habilidades são responsáveis por uma parcela significativa da variação encontrada em todo o sucesso socioeconômico do indivíduo. De acordo com os autores, a junção destas três literaturas permite desenvolver um modelo fiel para explicar seis fatos estilizados:

- i. As diferenças de habilidades entre os indivíduos e entre os grupos socioeconômicos abrem espaços para políticas em idades precoces. Uma implicação desta evidência empírica decorre que a qualidade das escolas e dos recursos educacionais tem relativamente pequenos efeitos sobre os déficits de habilidades e tem pouco efeito sobre os resultados dos testes das crianças de diferentes grupos socioeconômicos;
- ii. Há períodos sensíveis e críticos do desenvolvimento da criança. Assim, algumas habilidades são mais facilmente adquiridas em certas fases da infância do que as outras características;
- iii. Apesar dos baixos retornos para intervenções direcionadas para adolescentes carentes, a literatura empírica mostra altos retornos econômicos para investimentos corretivos em jovens crianças desfavorecidas;
- iv. Se o investimento precoce em crianças desfavorecidas não é acompanhado por investimentos mais tarde, o seu efeito em idades posteriores é menor.
- v. Efeitos de restrição de crédito sobre os resultados de adultos dependem da idade em que ele surgiu na família da criança. Ou seja, renda familiar durante os anos de faculdade desempenha um papel menor na determinação da participação do indivíduo.
- vi. Habilidades socioemocionais (não cognitivas) promovem habilidades cognitivas e são um importante produto de famílias afortunadas e intervenções bem sucedidas em famílias desfavorecidas.

Os seis fatos elencados tem fortes efeitos na condução de políticas públicas, principalmente em países com certo grau de subdesenvolvimento. Por exemplo, no Brasil, a

maior parte dos investimentos em educação é destinada a programas de incentivo ao acesso do ensino superior ou aqueles programas de formação profissional pública e de alfabetização de adultos. Em boa parte negligenciam-se os primeiros anos de vida dos indivíduos e posteriormente se gasta para tentar remediar as diferenças resultantes da falta de estímulos nos estágios iniciais do ciclo da vida.

Outro ponto que merece destaque neste trabalho, e que atualmente gera controvérsia entre acadêmicos e políticos, compreende o primeiro ponto, uma vez que os autores demonstram que, *a priori*, qualidade das escolas e dos recursos educacionais tem relativamente pequenos efeitos sobre os déficits de habilidades. Isto não quer dizer que qualidade educacional não seja importante, mas que investir em educação nos primeiros anos de vida são mais eficientes e efetivos na condução de uma política pública que preze pela igualdade de oportunidades.

Há ainda muito outros pontos relevantes na discussão apresentada por Cunha e Heckman (2007), e que aqui se pretende explorar no desenvolvimento deste trabalho. Portanto, configura-se o embasamento teórico aqui utilizado para investigar os efeitos da seca dos últimos anos no Semiárido Brasileiro sobre o desenvolvimento cognitivo dos indivíduos nascidos neste período. A hipótese é de que a restrição pluviométrica que assolou esta região pode ter comprometido tanto o estoque de saúde dos indivíduos quanto o desenvolvimento de suas habilidades e talvez possa explicar os baixos níveis educacionais desta região.

2.3 ESTRATÉGIA DE IDENTIFICAÇÃO EMPÍRICA

A estratégia empírica aqui desenvolvida permeia três aspectos. O primeiro tomar-se-á como base teórica o modelo desenvolvido por Cunha e Heckman (2007) de investimentos em *capacidades humanas*, aqui exposto e adaptado para redesenhar o problema a ser analisado. Segundo, apresenta-se o estimador de regressão quantílica com estrutura de dados em painel abordado em Canay (2011). Por fim, o terceiro ponto no qual se expõem os dados utilizados na construção do painel quantílico.

2.3.1 Modelo Teórico

A família desempenha um papel crucial na formação das habilidades cognitivas e não cognitivas, pois fornece tanto o *background* familiar, como as condições do meio ambiente com os quais tais habilidades são determinadas. Na busca de formalizar as evidências da literatura, Cunha e Heckman (2007) propõem um modelo de gerações sobrepostas de formação de habilidades no qual agrupa evidências da psicologia, economia e neurociência.

O modelo teórico proposto por Cunha e Heckman (2007) tem como base a teoria do ciclo da vida de formação de habilidades, em que apresenta como principal característica o desenvolvimento de um processo dinâmico de múltiplos estágios, ou seja, o ciclo de vida dos indivíduos é analisado considerando vários períodos e cada qual com uma taxa de retorno diferente. A ideia é mostrar que o desenvolvimento da criança depende de um conjunto de fatores. Assim, o modelo pode ser expresso como:

$$\theta_{t+1}^k = f_t^k(\theta_t^C, \theta_t^N, I_t^k, h^C, h^N), \quad k \in \{C, N\} \quad (1)$$

Onde θ_t denota o vetor de capacidades, ou seja, as habilidades cognitivas, habilidades não cognitivas: $\theta_t = (\theta_t^C, \theta_t^N)$. Deixe I_t denotar o vetor de investimentos cognitivos, não cognitivos e capacidades de saúde: $I_t = (I_t^C, I_t^N)$. Use $h = (h^C, h^N)$ para denotar habilidades parentais cognitivas e não cognitivas. A equação (1) traduz que em cada estágio t , pode-se definir uma tecnologia recursiva para habilidades cognitivas ($k = C$) e as habilidades não cognitivas, ($k = N$).

Em outras palavras, considera-se que os agentes possuem um vetor de *capacidades* em cada idade. Pressupõe-se que, no momento da concepção, a criança recebe condições iniciais genéticas¹⁹ e ambientais, θ_1 . Além disso, para cada investimento dos pais, I_t , gera um vetor de habilidades, e que as capacidades parentais, h (por exemplo: QI, genes, educação, renda, etc), geram *outputs* de investimentos de genes e de seus próprios pais. A tecnologia (1), na forma recursiva, quando o indivíduo é t anos de idade por ser escrita:

$$\theta_{t+1} = f_t(h, \theta_t, I_t) \quad (2)$$

¹⁹ Conforme Gluckman e Hanson (2005) e Rutter (2006), a expressão gene é desencadeada por condições ambientais iniciais.

para $t = 1, 2, \dots, T$. Assume-se que f_t é estritamente crescente e estritamente côncava em I_t , e duas vezes continuamente diferenciável para todos os argumentos. Substituindo em (2) por $\theta_t, \theta_{t-1}, \dots$, repetidamente, pode-se reescrever o estoque de recursos na fase $t + 1$, θ_{t+1} , em função de todos os investimentos do passado:

$$\theta_{t+1} = m_t(h, \theta_1, I_1, \dots, I_t), \quad t = 1, \dots, T \quad (3)$$

Conforme Heckman (2007) esta tecnologia é suficientemente rica para descrever a evidência de que a capacidade da criança de prestar mais atenção afeta o desempenho acadêmico posterior, pois crianças mais saudáveis são melhores alunos²⁰. Tal constatação é decorrente de duas características importantes da tecnologia: i) **autoprodutividade** – habilidades produzidas em um estágio aumentam as habilidades desenvolvidas nos estágios posteriores, e surge quando $\partial f_t(h, \theta_t, I_t) / \partial \theta_t > 0$; e ii) **complementariedade dinâmica** – habilidades produzidas em um estágio aumentam a produtividade do investimento em estágios subsequentes e surge quando $\partial^2 f_t(h, \theta_t, I_t) / \partial \theta_t \partial I_t' > 0$.

Além disso, os autores afirmam que esta tecnologia capta os períodos sensíveis e críticos postulados na literatura psicológica e da neurociência. Considere t^* ser o período sensível para θ_{t+1} , assim: $\frac{\partial \theta_{t+1}}{\partial I_s} = \frac{\partial m_t(h, \theta_1, I_1, \dots, I_t)}{\partial I_s} \equiv 0$ para todo $h, \theta_1, I_1, \dots, I_s, \quad s \neq t^*$

Mas, $\frac{\partial \theta_{t+1}}{\partial I_{t^*}} = \frac{\partial m_t(h, \theta_1, I_1, \dots, I_t)}{\partial I_{t^*}} > 0$ para alguns $h, \theta_1, I_1, \dots, I_t$

Esta condição diz que os investimentos em θ_{t+1} são produzidos no período t^* , mas não em qualquer outro período $s \neq t^*$. Período t^* é o período sensível para θ_{t+1} se:

$$\left. \frac{\partial \theta_{t+1}}{\partial I_s} \right|_{h=\bar{h}, \theta_1=\theta, I_1=i_1, \dots, I_t=i_t} < \left. \frac{\partial \theta_{t+1}}{\partial I_{t^*}} \right|_{h=\bar{h}, \theta_1=\theta, I_1=i_1, \dots, I_t=i_t}$$

Em outras palavras, o período de t^* é um período sensível em relação ao período de s se, ao mesmo nível de insumos, o investimento é mais produtivo na fase t^* do que em outro estágio $s \neq t^*$. A ideia é de que no **período sensível** haja certa estabilidade na conectividade dos neurônios decorrente da influência do ambiente, assim indivíduos nascidos em ambientes restritos tendem a produzir habilidades menos eficazes. Mesmo que se invistam períodos mais tardes, denominados de **períodos críticos**, este nunca será um substituto perfeito.

2.3.2 Identificação Empírica

O objetivo da pesquisa é investigar os efeitos das mudanças climáticas, corroborada pelos longos períodos de seca no nordeste, no desempenho dos jovens universitários do semiárido brasileiro. Assim, o estoque de *capacidades* do adulto $h' (= \theta_{T+1})$, é uma função de características dos pais, condições iniciais e investimentos durante a infância I_1 e I_2 :

²⁰ Ver Currie, 2008.

$$h' = m_2(h, \theta_1, I_1, I_2) \quad (4)$$

Dentro deste arcabouço, busca-se averiguar, em médio prazo, os efeitos das mudanças climáticas ocorridas nos últimos anos no semiárido brasileiro mais precisamente na Região Nordeste sobre a formação das habilidades ou *capacidades* dos indivíduos nascidos nesta região. Desta forma, considera-se um modelo simplificado de Cunha e Heckman (2007) com dois períodos ($t = 2$) seguidos por um período A quando adultos $\{t \in (T = 1, T + 2, T + A)\}$. O Resultado do adulto, y_{T+A} , é produzido por uma combinação de diferentes habilidades, no início do período $T = 1$ e $T + 2$:

$$y_{T+A} = f_t(\theta_t, \theta_P, I_t, \eta_{k,t}) \quad (5)$$

Seguindo a descrição de Cunha e Heckman (2010), as habilidades evoluem da seguinte forma. Cada indivíduo nasce com condições iniciais θ_1 . O primeiro período $T = 1$ compreende a fase intrauterina dos indivíduos, no qual sofrem influência dos ambientes iniciais (variável desvio das chuvas nos últimos 12 meses) e fatores genéticos (variáveis como raça e sexo), θ_t . Denota-se $h = \theta_P$ as habilidades parentais cognitivas e não cognitivas, como renda familiar, nível de escolaridade do pai e se exerce trabalho remunerado. Por sua vez, o período $T + 2$ compreende a fase antes da entrada na faculdade, sendo que I_t , gera um vetor de habilidades em $T + A$, I_2 . Neste caso, considera-se a nota de entrada no vestibular, a rede escolar em que o indivíduo estudou no ensino fundamental e médio.

E por fim, $\eta_{k,t}$, que denota choque e/ou *inputs* não observados que afetam o acúmulo de habilidade de cada indivíduo. Para não incorrer em problemas de viés, decorrente da existência de heterogeneidade não observada. Assume-se, ainda, que $\eta_{k,t}$ pode ser decomposta em dois componentes (π_k, ν_t) de tal modo que podemos reescrever a equação (5) pela tecnologia do painel:

$$y_{T+A} = y_{it} = \pi_k + \nu_t + \beta' x_{it} + u_{it} \quad (6)$$

Assume-se π_k como sendo um choque escalar independente sobre as pessoas, mas não ao longo do tempo. É um choque comum que afeta todas as tecnologias, mas o seu efeito pode ser diferente em cada uma das tecnologias. O componente ν_t , sua realização ocorre no final do período t , após as opções dos investimentos serem tomados e implementados. O pressuposto é de que as perturbações π_k e ν_t são escalares, embora as outras variáveis possam ser vetoriais e os distúrbios são independentes e identicamente distribuídos (*iid*). Para isso, apresenta-se o estimador de regressão quantílica com estrutura de dados em painel abordado em Canay (2011).

I – Estimador de Regressão Quantílica com efeitos fixos – QR(FE)

Canay (2011) apresenta um estimador consistente em que utiliza uma simples transformação de dados para eliminar os efeitos fixos dos *quantis* condicionais, conforme o modelo:

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it} \quad E(\varepsilon_{it}/X_{it}, \alpha_i) = 0 \quad (7)$$

Onde $(y_{it}, x_{it}) \in \mathbb{R} \times \mathbb{R}^k$ são variáveis observáveis, $\theta(U_{it}) = \beta$ e $(U_{it}, \alpha_{it}) \in \mathbb{R} \times \mathbb{R}$ são não observados. $U_{it}/x_{it}, \alpha_{it} \approx U[0,1]$. A função $\tau \rightarrow x'\theta(\tau)$ é assumida como sendo estritamente crescente em $\tau \in (0,1)$ e o parâmetro de interesse é assumido para ser $\theta(\tau)$. A principal restrição é que as heterogeneidades individuais associadas a x_{it} deve ter um efeito local puro.

Em suma, o método versa na questão em que as condições (U_{it}, α_{it}) do parâmetro $\theta(\tau)$ podem ser identificadas e consistentemente calculadas a partir dos dados. Para estimar a equação (7), Canay (2011) propõe um estimador simples de dois passos, sendo $E_T(\cdot) \equiv T^{-1} \sum_{t=1}^T(\cdot)$ e $E_{nT}(\cdot) \equiv (nT)^{-1} \sum_{t=1}^T \sum_{i=1}^n(\cdot)$, como segue:

Passo 1. Deixe $\hat{\theta}_\mu$ para $\sqrt{nT}$ – estimador consistente de θ_μ . Estimam-se os efeitos fixos como:

$$\hat{\alpha}_i \equiv E_T[y_{it} - x'_{it}\hat{\theta}_\mu] \quad (8)$$

Passo 2. Sendo $\hat{y}_{it} \equiv y_{it} - \hat{\alpha}_i$ e definir o estimar do dois-estágios $\hat{\theta}(\tau)$ como:

$$\hat{\theta}(\tau) \equiv \operatorname{argmin}_{\theta \in (\cdot)} E_{nT}[\rho_r(\hat{y}_{it} - x'_{it}\theta)] \quad (9)$$

Grosso modo, a ideia do método consiste em apurar o efeito fixo da variável dependente através de um estimador de média condicional (OLS), e depois aplicar uma regressão quantílica em dados em painel substituindo a variável dependente pelo valor deduzido do efeito individual.

2.3.3 Descrição dos Dados

Com o intuito de medir os impactos das secas no semiárido brasileiro durante a década de 90, ou seja, no período intrauterino dos jovens brasileiros que nasceram nesta região, sobre o estoque de *capacidades*²¹ dos indivíduos que prestaram vestibular de 2001 à 2010, a

²¹ Heckman (2007).

pesquisa norteou-se em três bancos de dados principais: i) Núcleo de Tecnologia da Informação (NTI); ii) Comissão Permanente de Vestibular (COPERVE); e iii) *Terrestrial Precipitation: 1900-2010 Gridded Montly Times Series* (versão 3.2).

O primeiro gerencia informações a respeito da trajetória acadêmica dos indivíduos e por isso possui um registro único de informações do desempenho acadêmico. O segundo é responsável pelo processo seletivo de alunos ingressantes nos cursos de graduação da UFPB e possui em seus registros o questionário socioeconômico dos alunos e sua respectiva nota de entrada nos cursos. Terceiro, tem-se os dados de precipitação utilizados nesta pesquisa. Tais bancos de dados possibilita a construção das variáveis requeridas para atender ao objetivo deste ensaio. Desta feita, lança-se mão de três blocos de variáveis de acordo com os períodos analisados na pesquisa. O quadro 2.1 resume as variáveis do estudo:

Quadro 2.1 – Descrição das Variáveis.

Tipo		Variável	
Bloco 1 (T=1)	θ_1	Clima	Desvio percentual da precipitação média.
	θ_1	Raça	<i>Dummy</i> para cor, assume-se 1 para brancos e 0 para não brancos
		Sexo	<i>Dummy</i> para sexo, assume-se 1 para masculino e 0 para feminino
Bloco 2 (T+2)	θ_p	Renda	Renda Familiar
		Escolaridade _{pai}	Nível de Escolaridade do Pai
		Ocupação _{mae}	<i>Dummy</i> se o pai ou a mãe trabalha, assume-se 1 para o exercício de trabalho remunerado e 0 o contrário.
	I_1	Rede _{fundamental}	<i>Dummy</i> para tipo de escola, 1 para pública e 0 para privada no ensino fundamental
		Rede _{médio}	<i>Dummy</i> para tipo de escola, 1 para pública e 0 para privada no ensino médio
Bloco 3 (T+A)	Y_{T+A}	Nota _{vestibular}	Nota de entrada no vestibular
		Nota _{disciplina}	Desempenho Acadêmico dos estudantes

Fonte: Elaboração Própria.

Das variáveis apresentadas, cabe ressaltar que as que compõem o Bloco 1 e 2, exceto clima, são comumente utilizadas na literatura de educação, sobretudo na de oportunidades educacionais (Ver Roemer, 1998; Bourguignon et. al., 2007; Ferreira e Gignoux, 2008; Gamboa e Waltenberg, 2012; entre outros). Contudo, o destaque da análise aqui referida encontra-se nas variáveis relacionadas aos fatores climáticos. Parte-se do pressuposto de que as flutuações de chuvas durante o período gestacional têm influência nos resultados dos indivíduos ao nascer e ao longo da vida. Ou seja, ter nascido em regiões com períodos restritivos de chuvas pode comprometer o estoque habilidades dos indivíduos.

Conforme averiguado no primeiro capítulo desta pesquisa, a ideia por trás é simples, aumentos de flutuações pluviométricas implicam em melhores condições para a produção agrícola e, conseqüentemente, maior disponibilidade de alimentos e nutrientes (Suliano et. al.

2009), bem como aumentam o acesso a água potável e redução de incidência de doenças infecciosas (Parry et al., 2009 - IPCC). Para captar tais fatos, semelhante a análise anterior, segue-se a construção da variável log desvio das chuvas nos últimos 12 meses, obtida:

$$P_{m\tau} = \ln \left(\sum_{t=\tau-11}^{\tau} p_{mt} \right) - \ln (\bar{p}_m) \quad (10)$$

onde p_{mt} indica a precipitação mensal no município m e mês t , $\bar{p}_m$ é a precipitação anual média histórica no município m , e τ indica o mês de nascimento de um indivíduo. Neste caso, a média histórica será calculada para cada município no período de 1961-1990.

2.4 RESULTADOS

Nesta seção apresentam-se os resultados da Função de Produção Educacional incorporada pelos efeitos das chuvas no período intrauterino dos indivíduos que nasceram no semiárido brasileiro sobre seu desempenho na universidade, com base no modelo de Cunha e Heckman (2007). Inicialmente, expõem-se algumas estatísticas descritivas das variáveis selecionadas para este estudo exibidas na Tabela 2.1. Posteriormente, estima-se a equação (6) a fim de obter inferências da relação causal comparando o método de Canay (2011) com estimadores de média condicional padrão.

2.4.1 Estatística Descritivas

A pesquisa consiste em selecionar os indivíduos oriundos do semiárido brasileiro que realizaram o exame de entrada do vestibular e obtiveram êxito com o ingresso nos cursos de graduação da UFPB no período de 2001 à 2010. O painel *aluno-disciplina-período* e painel *município-ano-nascimento*, possibilitados pelos registros administrativos gerenciados pelo NTI e os dados meteorológicos nos primeiros anos de nascimento, respectivamente, são interligados através da variável data de nascimento dos indivíduos. Além disso, tem-se um conjunto de dados que caracterizam os indivíduos e se consolida em um painel com 47.157 observações. O sumário dos dados é apresentado na Tabela 2.1:

Tabela 2.1 – Estatística Descritiva

Variáveis	Número de Obs.	Média	Desvio Padrão	Mínimo	Máximo
Desempenho acadêmico	47.157	6.65	3.2326	0	10
Precipitação	47.157	-0.007	0.3117	-1.622	0.8576
Raça	47.157	0.4436	0.4968	0	1
Sexo	47.157	0.559	0.4965	0	1
Renda Familiar	47.056	1.979	0.8048	1	3
Nível de escolaridade do pai	47.157	2.457	1.1668	1	4
Pai Trabalha	40.596	0.7181	0.4499	0	1
Mãe Trabalha	44.783	0.78009	0.4142	0	1
Rede Escolar (fundamental)	46.574	0.4204	0.4936	0	1
Rede Escolar (ensino médio)	47.103	0.6978	0.4592	0	1
Nota de entrada (vestibular)	46.890	545	68.899	390.2	752

Fonte: Elaboração Própria a partir dos dados da pesquisa.

A Tabela 2.1 apresenta a média, desvio padrão, valor mínimo e valor máximo das variáveis selecionadas neste estudo. Averigua-se que a nota média dos alunos nas disciplinas cursadas é em torno de 6,7, dos quais boa parte dos alunos é do sexo masculino (56%), são considerados brancos (55,63%) e possuem renda familiar em torno de até 6 salários mínimos. Além disso, constata-se que são filhos de pais em que a maioria trabalha tanto o pai, quanto a mãe, e também, tem nível de escolaridade baixa, no qual tem pai analfabeto (31,79%) e com ensino fundamental (14,57%).

No que tange os investimentos, a grande maioria estudou no ensino fundamental em escola privada (57,96%), denotando investimentos nos primeiros anos de ensino, no entanto, esses investimentos não foram acompanhados no ensino médio, uma vez que a maioria estudou no ensino público (69,79%). Pelo teste de entrada, verifica-se que a média é de 545 pontos de um total de 1.000, sendo que o valor máximo da amostra denota 752 pontos. Por fim, verifica-se que nesta amostra a incidência média de chuvas em um período de 12 meses é cerca de 108 milímetros, com uma precipitação log-desvio médio de -0,007 e 9,4% das observações correspondem as secas.

2.4.2 Efeito das Secas no desempenho acadêmico

A análise aqui requerida parte do pressuposto de que as chuvas em um ano de nascimento têm uma relação positiva com o desenvolvimento das habilidades cognitivas e não cognitivas dos indivíduos (Heckman, 2007; Cunha e Heckman, 2007). Isso porque os municípios pertencentes às áreas geográficas do semiárido são caracterizados pela irregularidade das chuvas. Restrições pluviométricas podem propiciar ambientes iniciais desfavoráveis, uma vez que há efeito negativo na cadeia produtiva, na saúde e na qualidade de vida da região (Cirilo, 2008).

Desta feita, investiga-se a relação causal entre precipitação no início da vida e os resultados educacionais dos indivíduos quando adultos. Para isso, construiu-se um painel *aluno-disciplina-ano*, no qual se acompanha as notas das disciplinas cursadas de um grupo de indivíduos que nasceram no semiárido brasileiro, amparadas por um conjunto de variáveis relacionadas as características dos indivíduos e de seus pais. A Tabela 2.2 apresenta tais estimações sob a ótica de três estimadores: *i*) Estimador de dados em painel com efeitos fixos (FE); *ii*) Estimador de Regressão Quantílica (QR) e; *iii*) Estimador de Regressão Quantílica com Efeitos Fixos (QR(FE)). Além disso, para cada método utilizou-se a precipitação dos 13 a 24 meses antes do nascimento e a precipitação dos 12 meses posteriores ao nascimento dos indivíduos como controle.

Na primeira linha das colunas (1), (2) e (3) tem-se que o resultado da variável de interesse desse estudo. A coluna (1) refere-se ao resultado do estimador de Dados em Painel com efeitos fixos, no qual se verifica uma correlação positiva e estatisticamente significativa entre chuvas e desempenho acadêmico na magnitude de 0,2838. O método contabiliza a heterogeneidade não observada invariante no tempo considerando-o que na média o efeito fixo é o mesmo para todos os indivíduos, entretanto, não permite conhecer distribuição condicional dos resultados. Já a coluna (2) contempla o estimador de regressão quantílica padrão no qual a estimativa foi estatisticamente insignificante. O método permite caracterizar toda a distribuição condicional de uma variável resposta e apesar de ser robusta a *outliers*, capta as diferenças apenas entre os *quantis*.

Tabela 2.2 – Resultados das estimações

Variável/Método	FE (1)	QR (2)	QR(FE) (3)
Chuvas antes do nascimento	0.2838*** (0.0714)	-0.0151 (0.0367)	0.0249*** (0.0303)
Raça	-0.2632*** (0.0423)	-0.0677** (0.0223)	-0.2159*** (0.0187)
Sexo	-0.9885*** (0.0829)	-0.6744*** (0.0225)	-0.9156*** (0.0182)
Renda familiar	-0.2387*** (0.0308)	-0.1162*** (0.0163)	-0.2141*** (0.0152)
Escolaridade do pai	-0.2177*** (0.0231)	-0.0973*** (0.0116)	-0.1672*** (0.0089)
Pai trabalha	0.4758*** (0.0848)	0.2281*** (0.0805)	0.4492*** (0.0259)
Mãe trabalha	0.0647 (0.0859)	-0.0694** (0.0297)	0.0833*** (0.0263)
Rede Escolar (fundamental)	-0.0875*** (0.0637)	0.0283 (0.0281)	0.0632*** (0.0227)
Rede Escolar (médio)	-0.0163*** (0.1275)	-0.0532*** (0.0312)	-0.1121*** (0.0262)
Nota de entrada	0.0116*** (0.0068)	0.0065*** (0.0002)	0.0398*** (0.0002)
Chuvas antes e após o nascimento	Sim	Sim	Sim
Observações	29.372	38.253	38.253

Fonte: Elaboração Própria a partir dos dados da pesquisa.

Nota: Erros padrão robustos entre parênteses. ***p-valor < 0,01. ** p-valor < 0,05. * p-valor < 0,10

Neste arcabouço de limitações e vantagens, aplicou-se o método de Canay (2011) que seja um estimador que une os dois métodos, no qual capta a heterogeneidade não observada que afeta a variável dependente, como também as características não observadas associadas as covariáveis ao longo da distribuição condicional. Neste sentido, a coluna (3) apresenta estimativa de 0,0249 estatisticamente significativa no intervalo de confiança de 99%. Este resultado implica em quatro desdobramentos. Primeiro, ao comparar-se a coluna (1) e (3), verifica-se que o viés da coluna (1) superestima o efeito das chuvas no período intrauterino dos indivíduos sobre o resultado educacional.

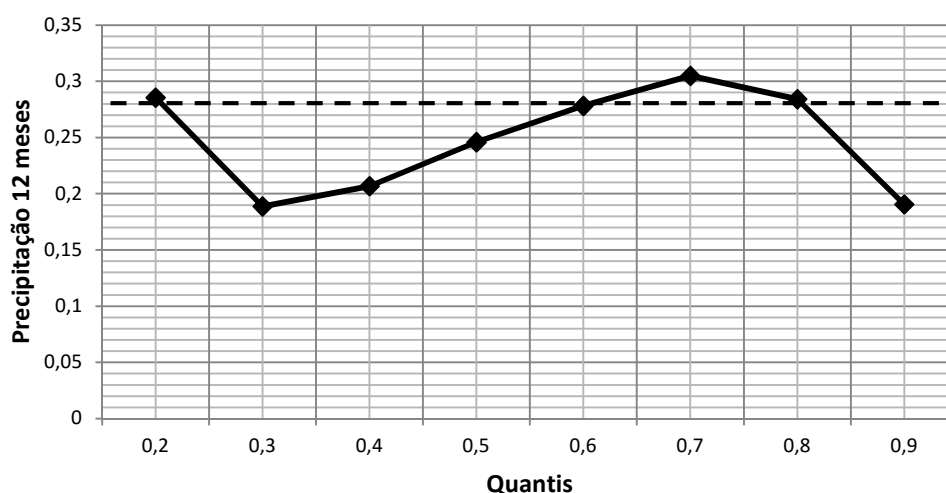
Segundo, tem-se que o resultado sugere que aumento de precipitação durante o período gestacional estão associados a aumentos no desempenho acadêmico universitário. Em outras palavras, na mediana, aumentos de chuvas na região semiárida tende a aumentar cerca de 2,5% o desempenho mediano dos estudantes. Tais constatações corroboram com a literatura vigente (Gluckman e Handson, 2005; Heckman, 2007; Stixrud e Urzua, 2006; Cunha e Heckman, 2007).

O terceiro desdobramento revela informações a respeito da covariáveis relacionadas aos investimentos que os pais realizam antes da entrada nos cursos universitários. Neste contexto, verifica-se que a nota de entrada tem efeito positivo implicando que quanto maior a nota no vestibular, melhor o desempenho ao longo do curso. Além disso, tem-se que alunos que estudam no ensino fundamental em escola pública tem efeito positivo. Parece que mesmo com a qualidade de educação menor em escola pública, o acompanhamento dos pais nesse período pode compensar tais implicações. Por outro lado, quando o indivíduo entra no ensino médio, outras variáveis como trabalho, desinteresse, entre outros podem comprometer o

desenvolvimento das suas habilidades, e uma escola pública tem efeito negativo no resultado educacional. Parece que a os investimentos em qualidade de educação do ensino médio é determinante para o seu desempenho.

O quarto desdobramento diz respeito aos efeitos ao longo da distribuição condicional, uma vez que o método consente conhecê-la. O Gráfico 2.1 apresenta tais estimativas com o estimador QR(FE). Entretanto, cabe aqui ressaltar que foram realizadas estimações com o estimador RQ para todos os *quantis*, porém todas as estimativas foram estatisticamente insignificantes, portanto, suprimidas do Gráfico em questão.

Gráfico 2.1 – Evolução por *quantis* dos coeficientes estimados por QR(FE)



Fonte: Elaboração própria a partir das estimações.

Graficamente, confirma-se o padrão qualitativo do efeito das chuvas: positivo e estatisticamente significativo a 1% ao longo da distribuição. A linha tracejada corresponde a estimativa média já apresentada na Tabela 2.2 – Coluna (1), que corresponde a um efeito de 0,2838. Note que esse resultado superestima o efeito da disponibilidade de água nos período da gravidez sobre resultado educacional principalmente nos *quantis* inferiores. No geral, constata-se que o impacto da precipitação média dos últimos 12 meses antes do nascimento cresce positivamente na medida em que cresce a desempenho acadêmico dos indivíduos. Além disso, observam-se disparidades entre os extremos. Alunos com melhores desempenhos, cauda superior, tem-se efeito menor das chuvas no seu desempenho, enquanto que na cauda inferior esse efeito é maior.

Em síntese, as evidências aqui colocadas complementam a literatura da epigenética, da neurociência e da psicologia, ao inferir que a baixa cognição dos indivíduos pode também ser produto da ausência de condições favoráveis durante o período intrauterino (Currie, 2008). As condições iniciais dos indivíduos tem papel relevante ao longo de toda vida (Heckman, Sticrud e Urzua, 2006; Heckman, 2007). E servem de respaldo para a aplicação de políticas públicas mais adequadas. Por exemplo, Cunha e Heckman (2007) advogam para política em

idades prematuras, uma vez que formação das habilidades são adquiridas com mais facilidades em idades precoces.

De acordo com os autores, há baixos retornos para intervenções direcionadas para adolescentes carentes, enquanto que a literatura empírica mostra altos retornos econômicos em intervenções em crianças desfavorecidas. Talvez políticas de acesso ao ensino superior tenha retorno econômico muito baixo e pode se refletir no baixo desempenho acadêmico dos indivíduos.

2.5 CONSIDERAÇÕES FINAIS

Este artigo apresenta evidências de uma relação causal entre as disponibilidades de água pluvial no semiárido brasileiro durante o período gestacional e o resultado econômico educacional. O cenário para o exercício empírico se deve ao painel quantílico *aluno-disciplina-ano* dos alunos que nasceram no semiárido e ingressaram nos cursos de graduação da UFPB no período de 2001 à 2010. Alicerçado pelo modelo teórico de Cunha e Heckman (2007) para a tecnologia de formação de habilidades, este trabalho acrescenta a utilização de regressão por *quantis* para uma estrutura de dados em painel, controlando para efeitos fixos.

Os resultados sugerem que quanto maior a precipitação durante o período gestacional dos indivíduos nascidos no semiárido maior é o desempenho acadêmico dos mesmos ao longo dos cursos de graduação da UFPB. Tais constatações corroboram com a literatura que se dedica aos estudos do impacto dos primeiros anos de vida dos indivíduos ao longo de sua vida (Gluckman e Handson, 2005; Heckman, 2007; Stixrud e Urzua, 2006; Cunha e Heckman, 2007). Além disso, averiguou-se que esse efeito positivo é persistente ao longo da distribuição condicional de desempenho, embora, varie em intensidades dos mesmos, principalmente nas caudas da distribuição.

Não obstante, os resultados encontrados abrem caminho para outras questões relacionadas aos determinantes do desempenho acadêmico dos universitários brasileiros, como o papel de políticas de acesso ao ensino superior, sistema de cotas, efeitos do tamanho da turma, entre outros. Ademais, destaca-se a importância desta pesquisa, como contribuição para os formuladores de políticas públicas, uma vez que a condição inicial, notadamente no período intrauterino e nos primeiros anos de vida, pode interferir na formação da cognição dos indivíduos e persistir ao longo da vida.

6 REFERÊNCIAS BIBLIOGRÁFICAS

Barker, D. J. P. Mothers, Babies and Health in Later Life. Editora Churchill Livingstone, edição nº 2, ISBN-10: 0443061653, 1998.

Barro, Robert J.; Lee, Jong Wha. A new data set of educational attainment in the world, 1950–2010. **Journal of development economics**, v. 104, p. 184-198, 2013.

BARROS, Ricardo Paes et al. Determinantes do desempenho educacional no Brasil. **Pesquisa e Planejamento Econômico**, v. 31, n. 1, 2001.

Becker, Gary S. Investment in human capital: A theoretical analysis. **The journal of political economy**, p. 9-49, 1964.

Becker, Gary S.; Tomes, Nigel. An equilibrium theory of the distribution of income and intergenerational mobility. **The Journal of Political Economy**, p. 1153-1189, 1979.

Buttner, Bettina; Thiel, Hendrik; Thomsen, Stephan Lothar. Variation of learning intensity in late adolescence and the impact on noncognitive skills. **ZEW-Centre for European Economic Research Discussion Paper**, n. 11-007, 2011.

Canay, Ivan A. A simple approach to quantile regression for panel data. **The Econometrics Journal**, v. 14, n. 3, p. 368-386, 2011.

Carneiro, Pedro; Heckman, James. **Human Capital Policy**. Institute for the Study of Labor (IZA), 2003.

Cirilo, José Almir. Políticas públicas de recursos hídricos para o semi-árido. **Estudos Avançados**, v. 22, n. 63, p. 61-82, 2008.

Checchi, Daniele; Dardanoni, Valentino. **Mobility comparisons: Does using different measures matter?**. Università di Milano, Dipartimento di economia politica e aziendale, 2002.

Cunha, Flavio; Heckman, James. The economics of human development: The technology of skill formation. **The American economic review**, v. 97, n. 2, p. 31-47, 2007.

Currie, Janet M. **The invisible safety net: Protecting the nation's poor children and families**. Princeton University Press, 2008.

Ferreira, Sergio Guimarães; Veloso, Fernando A. Mobilidade intergeracional de educação no Brasil. **Pesquisa e Planejamento Econômico**, v. 33, n. 3, 2003.

Fiori, N. As neurociências cognitivas. Tradução de Sonia M. S. Fuhrmann. Petrópolis RJ: Vozes, 2008.

Gluckman, Peter D.; Hanson, Mark A. **The fetal matrix: evolution, development and disease**. Cambridge University Press, 2005.

Hanushek, Eric A. The production of education, teacher quality and efficiency. In U.S. Office of Education, *Do Teachers Make a Difference?*, pages 79–99. Government Printing Office, Washington, DC, 1970.

Hanushek, Eric A. Conceptual and empirical issues in the estimation of educational production functions. **Journal of human Resources**, p. 351-388, 1979.

Heckman, James J. Lessons from the bell curve. **Journal of Political Economy**, p. 1091-1120, 1995.

Heckman, James. The technology and neuroscience of skill formation. **Invest in Kids Working Group, Center for Economic Development, Partnership for America's Economic Success**, July, v. 17, 2006.

Heckman, James J. The economics, technology, and neuroscience of human capability formation. **Proceedings of the national Academy of Sciences**, v. 104, n. 33, p. 13250-13255, 2007.

Lam, David. Generating Extreme Inequality: Schooling, Earnings, and Intergenerational Transmission of Human Capital in South Africa and Brazil. Research Report. 1999.

Lucas, Robert E. On the mechanics of economic development. **Journal of monetary economics**, v. 22, n. 1, p. 3-42, 1988.

Mankiw, N. Gregory; Romer, David; Weil, David N. A Contribution to the Empirics of Economic Growth. **The Quarterly Journal of Economics**, v. 107, n. 2, p. 407-437, 1992.

Mincer, Jacob. Investment in human capital and personal income distribution. **The journal of political economy**, p. 281-302, 1958.

Mincer, Jacob A. Schooling and earnings. In: **Schooling, experience, and earnings**. NBER, 1974. p. 41-63.

Piaget, Jean. **The origins of intelligence in children**. New York: International Universities Press, 1952.

Rebelo, Sergio. Long-Run Policy Analysis and Long-Run Growth. **Journal of Political Economy**, v. 99, n. 3, p. 500-521, 1991.

Rocha, Rudi; Soares, Rodrigo R. Water scarcity and birth outcomes in the Brazilian semiarid. **Journal of Development Economics**, v. 112, p. 72-91, 2015.

Romer, Paul M. Increasing returns and long-run growth. **The journal of political economy**, p. 1002-1037, 1986.

Romer, Paul M. Human capital and growth: theory and evidence. In: **Carnegie-Rochester Conference series on public policy**. North-Holland, 1990. p. 251-286.

Schultz, Theodore W. Investment in human capital. **The American economic review**, v. 51, n. 1, p. 1-17, 1961.

Schultz, Theodore W. The value of the ability to deal with disequilibria. **Journal of economic literature**, v. 13, n. 3, p. 827-846, 1975.

Weiss, C. The effects of cognitive and noncognitive abilities on earnings: different school systems. **European University Institute, unpublished**, 2010.

CAPÍTULO 3: Escassez da Água e Mortalidade Infantil: Novas Evidências a partir de um Painel Quantílico Robusto²²

Resumo

Propõe-se investigar os resultados de saúde expressos na taxa de mortalidade infantil para um painel de município-a-mês (1996-2010) do semiárido brasileiro, levando em consideração importantes efeitos: heterogeneidade não observada, a endogeneidade existente, bem como a confiabilidade dos dados de precipitação, até o momento negligenciado pela literatura. Os resultados, em conformidade com a literatura, mostram um efeito negativo da precipitação e um efeito positivo da seca sobre a taxa de mortalidade infantil dos municípios pertencentes ao semiárido brasileiro. Contudo, ressalta-se que os efeitos de tratamento aplicados ratifica o efeito negativo, no entanto, em magnitude inferior ao apresentado na literatura.

Palavras-chaves: Taxa de Mortalidade Infantil; Precipitação; Regressão Quantílica com efeitos fixos; Semiárido Brasileiro.

Abstract

We propose to investigate health outcomes expressed in child mortality rate for a county-by-month panel (1996-2010) of the Brazilian semiarid, taking into account important effects: unobserved heterogeneity, the existing endogeneity and reliability of precipitation data so far neglected in the literature. The results, in accordance with the literature, show a negative effect of precipitation and a positive effect of drought on the child mortality rate in the municipalities belonging to the Brazilian semiarid region. However, it is noteworthy that the effects of treatment applied confirms the negative effect, however, to a lesser extent than shown in the literature.

Keywords: Infant mortality rate; Precipitation; Quantile regression with fixed effects; Brazilian semiarid.

²² Com coautoria de Wallace Patrick Farias.

3.1 INTRODUÇÃO

Há um grande número de pesquisas que investigam os impactos das condições iniciais sobre os determinantes dos resultados socioeconômicos dos indivíduos quando adultos (Barcker, 1998; Gluckman e Handson, 2005; Cunha e Heckman, 2007; Heckman, 2007). Dentro desse contexto, choques climáticos nos anos iniciais de vida, sobretudo relacionados às quantidades pluviométricas, tem desempenhado papel relevante nos estudos sobre o desenvolvimento humano (Maccini e Yang, 2006; Ugaz e Zanolini, 2011; Burgess et al., 2014).

Dentre as principais consequências, ressaltam-se o efeito negativo em: i) processos produtivos (Aguilar e Vicarelli, 2011; Bezabih, Di Falco e Mekonnen, 2014); ii) na saúde (Currie e Vogl, 2012; Currie et al., 2013; Currie e Rossim-Slter, 2013; Rocha e Soares, 2015); iii) bem como na acumulação de capital humano (Shah e Steinberg, 2013). O pressuposto básico é de que os choques climáticos propiciam ambientes iniciais (des) favoráveis, uma vez que há efeito negativo/positivo na cadeia produtiva, na saúde e na qualidade de vida da região (Cirilo, 2008).

Nesse ínterim, a questão pluviométrica vem ganhando destaque em estudos recentes. Por exemplo, Shah e Stenberg (2013) ressaltam que exposição a períodos de seca no início da vida tem efeitos prejudiciais nos salários, na saúde e na escolaridade dos indivíduos adultos da área rural da Índia. Burgess, et al. (2014) ratificam esses resultados para a Índia em termos de mortalidade em populações rurais e urbanas, bem como altas temperaturas reduzindo salários agrícolas e reais. No Brasil, Rocha e Soares (2015) enfatizam a escassez da água na região semiárida do nordeste como fator determinante para a mortalidade infantil mais elevada, menor peso ao nascer, e períodos de gestação mais curtos.

Por outro lado, Maccini e Yang (2008) investigam o efeito das condições climáticas na época do nascimento sobre a saúde, educação e resultados socioeconômicos dos adultos indonésios e concluem que uma maior precipitação no início da vida tem grandes efeitos positivos sobre os resultados de mulheres adultas, mas não para homens. Este conjunto de evidências vem atribuindo determinado grau de importância para as questões pluviométricas no desenvolvimento humano destas regiões, contudo, é relevante ressaltar que estas estimativas podem estar viesadas se não forem observadas questões como confiabilidade de dados de clima e a endogeneidade existente nessas relações.

Em que pese o desenvolvimento recente da associação de clima no desenvolvimento humano dos indivíduos, estes questionamentos merecem especial atenção para não incorrer em possíveis sub ou superestimações de resultados, ou ainda, em erros de especificação. Tais tratamentos são importantes na definição do correto impacto das mudanças climáticas sob resultados de saúde, crucial para os formuladores de políticas públicas. Deste modo, o objetivo desta pesquisa é adicionar elementos à literatura existente no intuito de reduzir incertezas e fornecer *insights* sobre os resultados de saúde dos indivíduos enfatizando três conjunturas.

Primeiro, considera-se que as estimativas do efeito médio da variável dependente podem não retratar a realidade da maioria dos municípios. A explicação para tal é a existência de heterogeneidade entre os municípios que a regressão clássica de dados de painel não capta. Potencialmente, modelos de regressão quantílica poderiam explicar a heterogeneidade não observada e efeitos de localização, e, concomitantemente a inclusão dos efeitos fixos dos dados em painel comportando, assim, um modelo de regressão com estimadores mais robustos. Nesse sentido, propõe-se averiguar as implicações de restrições pluviométricas sobre os períodos iniciais de vida dos indivíduos ao longo da distribuição utilizando o método de regressões quantílicas em dados de painel.

Segundo, pretende-se ampliar a análise. O ponto central é a omissão de variáveis, uma vez que muitas pesquisas vêm atribuindo ao fator clima especial ênfase, e por ora deixando de levar em conta outros fatores que podem afetar a mudança dos padrões de saúde que não seja clima. Por exemplo, a distribuição geográfica e a sazonalidade de doenças transmitidas podem alterar resultados de saúde e não necessariamente a causa disso estar nas mudanças climáticas. Para tratar esses problemas empregou-se o método de Harding e Lamarche (2009), que seja um estimador de regressão quantílica de um modelo de dados em painel com variáveis independentes endógenas, no qual permite que a variável endógena possa ser correlacionada com fatores não observados que afetam a variável resposta.

Para a consecução do método necessitar-se-á de um vetor que denote uma variável que afeta o nível de precipitação, mas não tem efeito direto sobre os resultados de saúde dos indivíduos. Pesquisas tem mostrado que as condições oceânicas, sobretudo a Temperatura da Superfície do Mar (TSM), apresentam-se como a principal variável física a influenciar as condições climáticas oriundas de dois fenômenos climáticos: El Niño²³ – Oscilação Sul (ENOS) e Dipolo Atlântico, tornando-se, portanto, o instrumento utilizado na pesquisa. Neste caso captam-se oscilações dos Oceanos Atlântico e Pacífico que afetam a ocorrência de chuvas no semiárido brasileiro.

Terceiro, e não menos importante, pretende-se averiguar a base de dados de clima, em cenário com dados observacionais²⁴, estes considerados na literatura meteorológica e climatológica como mais confiáveis, uma vez que se verifica que boa parte da literatura supracitada utiliza dados históricos de variações de chuvas obtidas por técnicas de reanálise²⁵ – chamados de dados modelados. Embora ambas as variáveis reproduzam a climatologia e as variabilidades das regiões de interesse, há um grande debate²⁶ acerca da precisão dos dados de reanálise em razão de sub/superestimações de intensidade de chuvas na média. Portanto, objetiva-se testar o impacto das chuvas nos primeiros anos de vida utilizando dados

²³ Ver Kayano, Rao e Moura (1988), Rao, Lima e Franchito (1993), Xavier et. al. (2003).

²⁴ Dados observacionais são oriundos de estações meteorológicas de superfície. De acordo com Vieira e Piculli (2009), há dois tipos de estações: i) convencionais – onde os dados são coletados e organizados por um observador; e ii) automáticas – a coleta é feita através de sensores.

²⁵ O processo de reanálise captam dados oriundos de satélites/estações, no qual acontece por duas etapas. Primeiro são interpolados através de modelos utilizados para prever tempo e, posteriormente, novos dados são interpolados para áreas em que não possuem dados disponíveis (Freitas, Franchito e Rao, 2010).

²⁶ Ver Guedes et. al., 2005; Pinto et. al., 2009; Freitas, Franchito e Rao, 2010

observacionais obtidos no Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) através de estimação de dados de painel com efeitos fixos.

Em síntese, a estratégia empírica adotada basear-se-á em três etapas. Primeiro, baseado em Canay (2011), estimar-se-á por regressão quantílica com dados em painel considerando a heterogeneidade não observada e os efeitos das covariáveis em ambientes com efeito fixo. Segundo, visando endogeneizar a quantidade de precipitações ocorridas no semiárido do Nordeste brasileiro, utilizar-se-á a Temperatura da Superfície do Mar (TSM), como instrumento em um modelo empírico baseado em Harding e Lamarche (2009). Por último, na terceira etapa, com a base dados obtidos de estações meteorológicas observacionais, realizar-se-á estimação para comparação entre as duas diferentes bases a fim de comprovar sub/superestimações dos resultados de Rocha e Soares (2015).

Além desta introdução, o restante do trabalho está organizado em mais quatro seções. A próxima seção apresenta a estratégia empírica, bem como a descrição dos dados e concepção das variáveis utilizadas nas estimações. Na terceira seção são discutidos os resultados apurados e a quarta parte reservada para as considerações finais.

3.2 METODOLOGIA E DADOS

Esta seção destina-se a apresentar a estratégia adotada neste estudo. Inicialmente, apresenta-se a contexto teórico e posteriormente o modelo empírico para avaliar o impacto das restrições pluviométricas nos primeiros anos de nascimento nos resultados de saúde dos indivíduos, nascidos no semiárido na década de 90. A seguir, discorre-se sobre dois métodos adotados. O primeiro, baseado em Canay (2011) visa apresentar modelo de regressão quantílica com dados em painel com efeitos fixos. Por sua vez, o segundo baseia-se em Harding e Lamarche (2009) numa abordagem de regressão quantílica para modelos de dados em painel com variáveis endógenas e efeitos individuais associados com as variáveis independentes. Por fim, expõem-se os dados utilizados, bem como a variável instrumental Temperatura da Superfície do Mar (TSM).

3.2.1 Contexto

O estudo da literatura da economia da saúde remonta a análise comportamental desenvolvido por Grossman (1972) que, análogo à literatura sobre o capital humano, considera a saúde como resultante de um processo de produção, portanto, uma variável de estoque. A partir deste escopo, desenvolvimentos recentes relacionados a literatura da psicologia, neurociência cognitiva, economia e epidemiologia tem destacado que a condição inicial depende das características genéticas, mas também das condições iniciais vivenciadas pelos indivíduos.

Nesse contexto, o modelo teórico de Heckman (2007) é aqui utilizado e adaptado para fins desta pesquisa, e tem como base a teoria do ciclo da vida, em que apresenta como principal característica o desenvolvimento de um processo dinâmico de múltiplos estágios, ou seja, o ciclo de vida dos indivíduos é analisado considerando vários períodos e cada qual com uma taxa de retorno diferente. A ideia é mostrar que o desenvolvimento da criança depende de um conjunto de fatores.

Conforme Heckman (2007) esta tecnologia é suficientemente rica para descrever a evidência de que a capacidade da criança em ser mais saudável afeta o resultados de saúde posteriores. Tal constatação é decorrente de duas características importantes da tecnologia: *i) autoproductividade* – habilidades produzidas em um estágio aumentam as habilidades desenvolvidas nos estágios posteriores, e surge quando $\partial f_t(h, \theta_t, I_t) / \partial \theta_t > 0$; e *ii) complementariedade dinâmica* – habilidades produzidas em um estágio aumentam a produtividade do investimento em estágios subsequentes e surge quando $\partial^2 f_t(h, \theta_t, I_t) / \partial \theta_t \partial I'_t > 0$.

Além disso, os autores afirmam que esta tecnologia capta os períodos sensíveis e críticos postulados na literatura psicológica e da neurociência. Em outras palavras, o período de t^* é um período sensível em relação ao período de s se, ao mesmo nível de insumos, o investimento é mais produtivo na fase t^* do que em outro estágio $s \neq t^*$. A ideia é de que no **período sensível** haja certa estabilidade na conectividade dos neurônios decorrente

da influência do ambiente, assim indivíduos nascidos em ambientes restritos tendem a produzir habilidades menos eficazes. Mesmo que se invistam períodos mais tardes, denominados de **períodos críticos**, este nunca será um substituto perfeito. Nesse contexto, o objetivo da pesquisa é investigar os efeitos das mudanças pluviométricas, corroboradas pelos longos períodos de seca no nordeste, nos resultados de saúde de jovens nascidos no semiárido brasileiro na década de 90.

3.2.2 Estratégia empírica

A. Modelo Empírico

O painel é composto de municípios por mês segundo a região semiárida nordestina, para o período de 1996-2010, construído com as variáveis de interesse. Assim, a análise dos impactos das variações pluviométricas na mortalidade infantil durante o período de gestação segue a especificação do painel de municípios por mês de nascimento abordado em Rocha e Soares (2015):

$$Y_{itm} = \alpha + \beta R_{itm} + \phi_{im} + \lambda_t + \varphi t_{gtm} + \pi T_{itm} + \epsilon_{itm} \quad (1)$$

onde Y_{itm} é um resultado de saúde médio (por município) para as crianças nascidas no município i , no ano t e mês m . A variável dependente utilizada é a mortalidade infantil. R_{itm} é a variável de precipitação; ϕ_{it} é um efeito-fixo para município i e mês m (com $m = 1, 2, \dots, 12$); λ_t é um efeito-fixo ano; T_{itm} é a temperatura média no município i no mesmo período de 12 meses antes do nascimento; t_{gtm} é a tendência temporal linear e ϵ_{itm} é o termo de erro aleatório.

Cabe aqui destacar que a variável relacionada ao clima, R_{itm} , é análoga ao realizada em Rocha e Soares (2015), ou seja, são concebidas de mesma forma, embora difiram na natureza do problema, uma vez que a fonte de dados e sua tabulação são diferentes entre si. Esta variável é construída de duas maneiras, ambas para captar os efeitos das secas no semiárido através da flutuação das chuvas durante o período de gestação dos indivíduos:

- I. **Log desvio das chuvas nos últimos 12 meses** – conforme os autores, esta variável por ser interpretada como uma porcentagem de desvio da precipitação média. Definida por:

$$R_{it}^1 = \ln \left(\sum_{t=\tau-11}^{\tau} r_{im} \right) - \ln (\bar{r}_i) \quad (2)$$

onde r_{im} indica a precipitação mensal no município i e mês m , $\bar{r}_i$ é a precipitação anual média histórica no município i , e τ indica o mês de nascimento de um indivíduo. Diferente

ao realizado em Rocha e Soares (2015), o cálculo do $\sum_{t=\tau-1}^{\tau} r_{im}$ será realizado apenas para os municípios pertencentes ao semiárido, uma vez que contabilizar estados onde tem muita chuva, por exemplo, os estado do Norte, pode superestimar esta covariável.

- II. **Seca dos últimos 12 meses** – uma variável *dummy* que assume o valor 1 se houve seca no período e 0 se o contrário. Em síntese esta variável busca captar os eventos extremos, calculado por:

$$R_{it}^2 = 1 \text{ se } \sum_{t=-11}^{\tau} r_{im} < (\bar{r}_i - r_i^{SD}), \text{ e } 0 \text{ para outros} \quad (3)$$

onde r_i^{SD} é o histórico anual do desvio padrão de chuvas para o município i (calculado sobre o período de 1961-1990). Em outras palavras, o $R_{it}^2 = 1$ indica que a precipitação ao longo dos 12 meses anteriores ao nascimento de um indivíduo era mais do que um desvio padrão abaixo da média histórica para o município i .

Em suma, a equação (1) estima os efeitos das variações pluviométricas, aqui representadas nas equações (2) e (3), sobre a taxa de mortalidade infantil em dados em painel de município-a-mês de 1996 a 2010.

B. Métodos Empíricos

Para estimar a equação (1) é necessário considerar que os efeitos fixos por município-mês controlam variações climáticas em determinados períodos do ano assim como diferenças entre os municípios, enquanto que o efeito fixo por ano tenta captar choques agregados e tendências que impactam toda a região. Por sua vez, a tendência linear controla a dinâmica da evolução das mudanças climáticas e socioeconômicas da região, e a temperatura é responsável por outras variações climáticas que podem estar relacionadas com a quantidade de chuvas (Rocha e Soares, 2015).

Uma grande literatura sobre estimação de regressões quantílicas com dados em painel tem surgido nos últimos anos, tal como Koenker (2004) que introduziu uma abordagem geral para estimação com dados longitudinais. Controlando para os efeitos fixos municipais e a heterogeneidade das covariáveis, a estimação por *quantis* é uma abordagem mais flexível para a análise de dados em painel do que os estimadores de efeitos fixos e aleatórios gaussianos clássicos. Em adição, considerar os efeitos de tratamento em modelos quantílicos de dados em painel constitui mais recente robustez das análises desta natureza. Não obstante, é necessário ir além e também captar outros efeitos na média. Portanto, neste arcabouço, apresentam-se os dois estimadores aqui utilizados.

I - Estimador de Regressão Quantílica com efeitos Fixos- Canay (2011) – (QRFE)

Canay (2011) apresenta um estimador consistente em que utiliza uma simples transformação de dados para eliminar os efeitos fixos dos *quantis* condicionais, conforme o modelo:

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it} \quad E(\varepsilon_{it}/X_{it}, \alpha_i) = 0 \quad (4)$$

Onde $(y_{it}, x_{it}) \in \mathbb{R} \times \mathbb{R}^k$ são variáveis observáveis, $\theta(U_{it}) = \beta$ e $(U_{it}, \alpha_{it}) \in \mathbb{R} \times \mathbb{R}$ são não observados. $U_{it}/x_{it}, \alpha_{it} \approx U[0,1]$. A função $\tau \rightarrow x'\theta(\tau)$ é assumida como sendo estritamente crescente em $\tau \in (0,1)$ e o parâmetro de interesse é assumido para ser $\theta(\tau)$. A principal restrição é que as heterogeneidades individuais associadas a x_{it} deve ter um efeito local puro.

Em suma, o método versa na questão em que as condições (U_{it}, α_{it}) do parâmetro $\theta(\tau)$ podem ser identificadas e consistentemente calculadas a partir dos dados. Para estimar a equação (7), Canay (2011) propõe um estimador simples de dois passos, sendo $E_T(\cdot) \equiv T^{-1} \sum_{t=1}^T(\cdot)$ e $E_{nT}(\cdot) \equiv (nT)^{-1} \sum_{t=1}^T \sum_{i=1}^n(\cdot)$, como segue:

Passo 1. Deixe $\hat{\theta}_\mu$ para $\sqrt{nT}$ – estimador consistente de θ_μ . Estimam-se os efeitos fixos como: $\hat{\alpha}_i \equiv E_T[y_{it} - x'_{it}\hat{\theta}_\mu]$.

Passo 2. Sendo $\hat{y}_{it} \equiv y_{it} - \hat{\alpha}_i$ e definir o estimar do dois-estágios $\hat{\theta}(\tau)$ como:

$$\hat{\theta}(\tau) \equiv \operatorname{argmin}_{\theta \in (\cdot)} E_{nT}[\rho_r(\hat{y}_{it} - x'_{it}\theta)]$$

O método consiste na obtenção de *insights* sobre os efeitos dos regressores na variável dependente. Isto porque permite uma leitura dos efeitos heterogêneos através da transformação dos dados, uma vez que em modelos de regressão padrão, que seguem a média, a heterogeneidade muitas vezes não é captada.

II - Estimador de regressões quantílicas de dados em painel com variáveis instrumentais (QRVI)

Alternativamente, considerar-se-á para efeitos de resultados, o estimador proposto por Harding e Lamarche (2009), no qual apresentam uma abordagem de regressão quantílica para modelos de dados em painel com variáveis endógenas. Este método é mais robusto do que o método de Canay (2011), uma vez que potencializa as estimativas ao considerar na regressão

efeitos além da heterogeneidade não observada. O método é uma extensão do procedimento de Chernozhukov e Hansen (2008) permitindo a inclusão de efeitos fixos introduzidos em Koenker (2004). Denote o vetor de variável endógena, d , na equação (4):

$$y_{it} = d'_{it}\delta + x'_{it}\beta + \alpha_i + u_{it} \quad (5)$$

$$d_{it} = h(x_{it}, w_{it}, v_{it}) \quad (6)$$

$$\alpha_i = g(x_{i1}, \dots, x_{iT}, d_{i1}, \dots, d_{iT}, \varepsilon_{it}) \quad (7)$$

onde a equação (6) define a variável endógena d relacionada a um vetor de instrumentos w que estão estocasticamente independente de u . Já a variável v é estocasticamente dependente de u . A equação (7) expressa a correlação entre as variáveis e os efeitos individuais. Neste caso, assume-se que ε é independente de u e v . Considere a função objetivo para o relacionamento quantílico instrumental condicional:

$$R(\tau, \delta, \beta, \gamma, \alpha, \mu) = \sum_{t=1}^T \sum_{i=1}^N \rho_{\tau}(y_{it} - d'_{it}\delta - x'_{it}\beta - \alpha_i - z'_{it}\alpha - \hat{w}'_{it}\gamma) \quad (8)$$

Onde $\hat{w}_{it}$ é um vetor de instrumentos e $\rho_{\tau} = u(\tau - 1(u \leq 0))$ é uma função perda de regressão quantílica. Seguindo o procedimento de dois passos de Chernozhukov e Hansen (2006, 2008), temos:

Passo 1. Minimizar a função objetivo (8) para β, α, μ e γ em função de τ e δ :

$$\{\hat{\beta}(\tau, \delta), \hat{\gamma}(\tau, \delta), \hat{\alpha}(\tau, \delta)\} = \arg \min_{\beta, \gamma, \alpha} R(\tau, \delta, \beta, \gamma, \alpha) \quad (9)$$

Passo 2. Em seguida, estima-se o coeficiente da variável endógena para encontrar o valor de δ , no qual minimiza a função distância ponderada definida em y :

$$\hat{\delta}(\tau) = \arg \min_{\delta} \hat{\gamma}(\tau, \delta)' A \hat{\gamma}(\tau, \delta) \quad (10)$$

Onde A é uma matriz definida positiva. De acordo com Harding e Lamarche (2009) este estimador é assintoticamente normal podendo ser executado para mais de um quantil simultaneamente.

3.2.3 Descrição dos dados

A. Dados de Clima

O uso de variáveis relacionadas ao clima é empregado para suportar uma ampla gama de estudos de investigação sobre o impacto destas variáveis sobre os resultados de saúde dos indivíduos. Neste sentido, o uso de dados de precipitação e temperatura vem ganhando espaço da literatura, pois a variabilidade destas afetam diversas atividades econômicas (Freitas, Franchito e Rao, 2010) e costumam representar a variabilidade de sinais geofísicos e regimes de precipitação (Guedes et. al., 2004), sobretudo na literatura climatológica. O Quadro 3.1 apresenta algumas fontes e tipo de análise.

Quadro 3.1 – Diversas Fontes e Metodologias de dados de Precipitação

Tipos de Dados	Fonte	Grade	Autores que utilizaram os dados
Observacionais	ANEEL	-	Freitas, Franchito e Rao, 2010;
	SUDENE	-	Menezes, et. al (2008)
	BDMEP	-	Rao et. al., 1996;
	DNOCS	-	Rao, Lima e Franchito (1993)
Reanálise	NCEP/NCAR	~1,9° x 1,9°	Guedes et. al., 2005; Pinto et. al., 2009; Freitas, Franchito e Rao, 2010;
	CPTEC	0,4° x 0,4°	Pinto et. al., 2009;
	CPC/NOOA		
	Terrestrial Air Temperature and Terrestrial Precipitation	0,5° x 0,5°	Rocha e Soares (2015)
Sensoriamento	GPCP	2,5° x 2,5°	Pinto et. al., 2009; Freitas, Franchito e Rao, 2010;
	CMAP	2,5° x 2,5°	Pinto et. al., 2009;
	TRMM	1,0° x 1,0°	Pinto et. al., 2009;

Fonte: Elaboração dos autores.

Na literatura meteorológica e climatológica, destacam-se três tipos de dados: i) dados observacionais - estações; ii) dados de reanálise e iii) dados de sensoriamento. O primeiro é oriundo de estações compostas por sensores isolados que registram continuamente os parâmetros meteorológicos que são lidos e anotados por um observador a cada intervalo. Contudo, apesar de reconhecer a fidedignidade deste tipo de observação, questões geográficas e econômicas tem dificultado a produção de séries temporárias de longo prazo (Pinto et. al., 2009).

Por sua vez, dados de reanálise são muito convenientes para completar as lacunas onde não possuem informações, como por exemplo, onde não há estação meteorológica, ou mesmo, devido a problema de obtenção de dados. No entanto, as séries históricas são geradas a partir de modelos e são construídas com base em algum tipo de observação em superfície, inclusive de estações. Por fim, os dados de sensoriamento remoto combinam dados de superfície com estimativas feitas por satélite para calibrar o banco de dados.

É possível perceber que existe um *trade-off* no uso de variáveis clima. Se por um lado, os dados observacionais são fontes mais confiáveis, a ausência de informações e as quebras das séries de longo prazo tem dificultado as investigações climatológicas. Todavia, os dados de reanálise dependem de fontes primárias confiáveis, modelos bem ajustados e técnicas de obtenção robustas. Neste sentido, Silva, *et. al.* (2007), considerando as novas calibrações efetuadas no modelo de obtenção dos dados de reanálise da CPC/NCAR, realizado em 2005, constataram através dos diagramas de dispersão tendências para valores mais altos da CPC em relação aos dados observacionais. Não obstante, os autores ainda ressaltam, consistentemente aos estudos anteriores, que as comparações indicam menor número de dias secos, maior número de dias de baixa precipitação, e menos eventos extremos de precipitação por grade analisada, apesar da boa correlação existente entre as mesmas.

Outros estudos também visam comparar alguns produtos de precipitação. Por exemplo, Freitas, Franchito e Rao (2010) que compararam os dados ANEEL (observacionais); GPCP (sensoriamento) e NCEP/NCAR (reanálise) e concluíram que os dados de reanálise NCEP/NCAR superestimam a precipitação no Nordeste Brasileiro e que os dados observacionais ANEEL, apesar da ausência de dados e falhas reproduziram melhor a climatologia. Guedes *et. al.* (2005) em um estudo para Fortaleza-CE compararam dados NCEP/NCAR com os dados das estações e concluem que os dados observados e de reanálise são comparáveis, mas o ciclo anual da reanálise parece estar um mês adiantado em relação aos dados observados e que os desvios padrões mensais apresentam grandes diferenças. Tais resultados são corroborados em Pinto *et. al.* (2009) ao concluir que os dados de reanálise CPTEC não são capazes de representar os principais regimes de precipitação existentes no continente, com forte tendência a superestimar a intensidade média no interior do Nordeste Brasileiro.

Nesse arcabouço, recentemente Rocha e Soares (2015) utilizaram os dados *Terrestrial Precipitation: 1900-2010 Gridded Montly Times Series* (versão 3.02). Este conjunto de séries temporais é procedente das estações do *Global Historical Climatology Network* (GHCN) e *Global Surface Summary of day* (GSOD) através de interpolação espacial. Portanto, trata-se de dados de reanálise e que de acordo com os estudos supracitados podem-se incorrer em superestimações de valores de precipitação e temperatura para o semiárido brasileiro. Para fins de comparação com os resultados recentes dos autores e em virtude da ausência de séries mais longas, neste artigo empregam-se os dados observacionais do Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP), obtidos através do Instituto Nacional de Meteorologia (INMET).

A ideia é verificar se os impactos das restrições pluviométricas na época de nascimento do indivíduo são realmente relevantes nos resultados de saúde quando adultos em cenário com dados mais confiáveis de precipitação, embora isso acarrete na redução da amostra. Isto porque este banco não possui estações para todos os municípios pertencentes do semiárido brasileiro como também os inícios das séries históricas divergem entre elas.

Neste escopo foram criadas as duas variáveis expressas nas equações (1) e (2) utilizando os dados obtidos na BDMEP, primeiro localizando o município através da latitude

e longitude; e depois, gerando os dados mensais de precipitação para cada município que compõe o semiárido brasileiro para o período de 1977 à 2010. No total, foram geradas 8.370 observações distribuídas nos 49 municípios analisados.

B. Variável Instrumental

Para empregar o método de Harding e Lamarche (2009), que seja um estimador de regressão quantílica de dados em painel com variáveis instrumentais necessitar-se-á de um vetor que denote uma variável que afeta o nível de precipitação, mas não tem efeito direto sobre os resultados de saúde dos indivíduos. Pesquisas tem mostrado que as condições oceânicas, sobretudo a Temperatura da Superfície do Mar (TSM), apresentam-se como a principal variável física a influenciar as condições climáticas.

Um fenômeno climático de interação oceano-atmosfera bastante investigado é o *El Niño* – Oscilação Sul (ENOS). Esta variabilidade climática é cíclica, oriunda do oceano Pacífico, com periodicidade que varia de 3 a 7 anos, duração igual ou superior a cinco meses consecutivos (Trenberth, 1997), e podem ser detectadas através das anomalias persistentes da TSM. Anos de *El Niño* (ou *El Niña*) estão normalmente associados à escassez (ou chuvas abundantes). A influência dos ENOS, sobretudo no Nordeste Brasileiro, tem sido objeto de estudo a destaca Kayano, Rao e Moura (1988); Rao, Lima e Franchito (1993); Xavier et. al. (2003); Alves e Souza (2003); Santos e Brito (2007), entre outros.

Outro fenômeno oceano-atmosfera é o dipolo do Atlântico. Destacam-se os estudos seminais de Hastenrath e Heller (1977), Moura e Slukla (1981), e mais recentemente, Aragão (1998), Moraes Neto e Araújo (2007), Andreoli e Kayano (2007), entre outros. Identifica-se o Dipolo quando as águas do Atlântico tropical estão mais quentes e do Atlântico Equatorial e Tropical Sul estão mais frias culminando para movimentos descendentes transportando o frio e seco inibindo a formação de nuvens e diminuindo a precipitação (Aragão, 1998).

Com base na literatura climatológica, optou-se pelos dados de temperatura dos Oceanos Pacíficos e Atlântico, conforme as pesquisas retromencionadas, uma vez que tanto os ENOS quanto o Dipolo Atlântico afetam a ocorrência de secas no semiárido brasileiro, mas não afetam diretamente a taxa de mortalidade infantil nem as condições de saúde ao nascer. Desta feita, os dados de TSM observados no Oceano Pacífico e Atlântico foram extraídos do *International Comprehensive Ocean–Atmosphere Dataset* (ICOADS) e *National Ocean and Atmospheric Administration* (NOAA), respectivamente.

O primeiro²⁷ arquivo contém dados mensais em pontos de grade de 2° x 2° de latitude-longitude para um período de 1854 – até o presente composto por quatro arquivos a seguir: i) Extremo Oriente - Niño 1 + 2 (0-10S; 90W-80W); ii) Leste Tropical - Niño 3 (5N-5S; 150W-90W); iii) Leste Central Tropical - Niño 3.4 (5N-5S; 170W-120W); e iv) Tropical Central – Niño 4 (5N-5S; 160E – 150W). Por sua vez, o segundo²⁸ arquivo são dados com resolução 0,25° x 0,25° obtidos através de uma interpolação ótima composto de dois tipos: i) Atlântico

²⁷ Ver detalhes são encontrados em Huang, et. al (2014) e Liu, et. al. (2014).

²⁸ Ver detalhes em Reynolds, et. al (2007).

Norte (5N-20N; 60W-30W); e ii) Atlântico Norte (0-20S; 30W-10E). A Tabela 3.1 apresenta os coeficientes de correlação de *Spearman* para os dados de precipitação e os dados de TSM.

Tabela 3.1 – Correlação de *Spearman* entre dados de Precipitação de TSM

DADOS PRECIPITAÇÃO/TSM	OCEANO PACÍFICO				OCEANO ATLÂNTICO	
	Niño 1 + 2	Niño 3	Niño 3.4	Niño 4	Atlântico Norte	Atlântico Sul
BMEP	0,229 (0,000)*	0,0047 (0,673)	0,239 (0,000)*	0,0098 (0,377)	-0,461 (0,000)*	0,469 (0,000)*
Rocha e Soares (2015)	0,211 (0,000)*	-0,0211 (0,056)	0,219 (0,000)*	-0,019 (0,089)	-0,480 (0,000)*	0,472 (0,000)*

Fonte: Elaboração dos Autores

* Estatisticamente significativo a um nível de significância de 5%.

Para fins desta pesquisa, utilizar-se-á os dados de TSM com maior correlação com os dados de precipitação. Logo, observa-se na Tabela 1 que o Niño 3.4 correspondente ao leste central tropical do Oceano Pacífico apresenta maior correlação com as duas bases de dados de precipitação e sendo o motivo pelo qual utilizamos na pesquisa.

3.3 RESULTADOS EMPÍRICOS

Para este estudo, os dados foram coletados a partir de áreas geográficas dos quais as chuvas são bastante irregulares e o solo é raso. O semiárido nordestino brasileiro possui essas características, e embora exista ocorrência de chuvas, historicamente são permeados por longos períodos de secas, com características edafoclimáticas e socioeconômicas restritivas à população em razão de sua escassez. O saldo global dos efeitos sobre a saúde é provável que seja negativo, principalmente quando acometem indivíduos no período intrauterino ou nos primeiros anos de nascimento (Backer, 1998; Cunha e Heckman, 2007; Heckman, 2007), onde tendem a ser particularmente vulneráveis aos efeitos adversos.

Dentro desse contexto, a análise da relação entre as condições externas nos anos iniciais de vida, especificamente em ambientes com escassez de água, sobre os resultados de saúde dos adultos – aqui expressos pela taxa de mortalidade infantil, é objeto desta seção dos quais se consideram três importantes inovações avaliadas a seguir.

3.3.1 Heterogeneidade não-observada

Ainda que de difícil identificação por não serem diretamente observáveis ou mensuráveis, a heterogeneidade não podem ser negligenciada nas estimações. Isso se deve ao fato de que como as estimativas se tratam de efeito médio, portanto, é plausível admitir que estas não retratem em sua totalidade a realidade dos 1.048 municípios analisados. Além do mais, a literatura epigenética (Fiori, 2008), na condução dos estudos de interações gene-ambiente, ratifica a clara distinção entre os atributos adquiridos ao longo da vida e aqueles relacionados aos genes, muitas vezes negligenciados na literatura da economia da saúde.

Assim sendo, características não observáveis correlacionadas com a probabilidade de exposição a ambientes desfavoráveis poderiam ser responsáveis por uma parcela dos resultados de saúde. Outrora, condições inerentes aos indivíduos são particularmente associadas aos seus resultados de saúde independente do ambiente em que convive. Não obstante, é relevante considerar, também, que os indivíduos podem responder a choques externos como a seca de forma distinta entre eles. Ademais, questões como distribuição geográfica e sazonalidade dever-se-iam também ser consideradas.

Uma alternativa para contornar estes problemas é o método de regressão quantílica com dados em painel proposto por Canay (2011) – denominado aqui de estimador QR(FE), apresentados na Tabela 3.2. Para obter as estimativas, cabe aqui enfatizar que fora utilizado as seguintes variáveis de controle: *i*) temperatura média dos últimos doze meses; *ii*) tendência temporal baseada na grade formada pelos municípios pertencentes ao quadrado 5° x 5° de latitude/longitude; *iii*) precipitação dos 13 a 24 meses antes do nascimento; *iv*) precipitação nos doze meses posteriores ao nascimento dos indivíduos. Não obstante, é relevante destacar que para cada painel construiu-se as variáveis de precipitação e seca, expressos nas equações (2) e (3) deste estudo e apresentados no Painel (A) e (B) das Tabelas, respectivamente.

Tabela 3.2 - Resultados das estimações com método de Canay (2011) do período de 1996-2010 da Região Semiárida do Nordeste

	FE			QR(FE)		
	(1)	(2)	(3)	(4)	(5)	(6)
Painel A – Precipitação						
Precipitação antes do nascimento	-3,25* (1,775)	-5,48*** (1,507)	-6,14*** (1,989)	1,09*** (0,129)	-2,08*** (0,134)	-3,02*** (0,141)
Precipitação 13-24 meses antes do nascimento			-1,49 (2,218)			-0,91*** (0,127)
Precipitação 1-12 meses após o nascimento			-2,74 (2,339)			-2,88*** (0,142)
Painel B – Seca						
Seca antes do nascimento	2,54 (1,801)	3,34*** (1,175)	3,73*** (1,227)	18,83*** (0,231)	2,67*** (0,147)	2,95*** (0,161)
Precipitação 13-24 meses antes do nascimento			-0,980 (2,154)			-1,15*** (0,125)
Precipitação 1-12 meses após o nascimento			-1,812 (2,176)			-2,69*** (0,132)
Observações	182.458	180.659	168.267	182.458	182.458	169.946
Número de Municípios	1.048	1.037	1.037	1.048	1.048	1.048

Fonte: Elaboração com base nas estimações.

Notas: Erros padrão robustos entre parênteses. ***p-valor < 0,01. ** p-valor < 0,05. * p-valor < 0,10. As colunas (7) e (10) não considera nas estimações a temperatura antes do nascimento.

A Tabela 3.2 apresenta os resultados utilizando o método de dados em painel com efeitos fixos (FE) expressos nas colunas (1), (2) e (3) e o estimador robusto de Canay (2011) de regressão quantílica com dados em painel (QRFE) apresentados nas colunas (4), (5) e (6), onde a variável dependente é a mortalidade infantil por cada 1000 nascimentos, mensurados por mês de nascimento em termos de nível municipal. O Painel A interpreta a precipitação como uma porcentagem da precipitação média, enquanto que o Painel B como uma *dummy* para seca.

Vislumbrando apenas o Painel A tem-se na Coluna 1 que, em termos de magnitude, isto representa que o aumento no nível de precipitação no período gestacional reduz em cerca de 3% a taxa de mortalidade infantil. Tais valores aumentam quando controlado pela temperatura média durante a gravidez (-5,48%) e nível de precipitação antes da concepção e depois do nascimento (-6,14%), conforme a Coluna (2) e (3), respectivamente.

Posteriormente, ao comparar-se os estimadores FE(RS) e QR(FE), cabem aqui alguns apontamentos. Primeiro, verifica-se na coluna (3) – FE(RS) - que os controles utilizados não foram estatisticamente significativos corroborando para a conclusão de Rocha e Soares (2015) de que níveis de chuva em outros períodos que não seja o gestacional não afetam a mortalidade infantil. O segundo, diz respeito a implicação do primeiro, uma vez que afirmar que as variáveis de controle não exercem influência sob a mortalidade infantil seria concluir que nutrição materna no ano antes da concepção, bem como a nutrição e doenças no primeiro ano de vida não são afetados pelas chuvas. Tal conclusão é justamente oposta a literatura vigente (Glewwe e King, 2001; Cunha e Heckman, 2007; Kudamatsu et. al., 2010).

Terceiro, tais evidências são refutadas com o estimador QR(FE), mostrando que as variáveis de controle são estatisticamente significativas e negativas e a principal implicação é a redução da magnitude da variável resposta. Além do mais, verifica-se através da coluna (6) que o efeito da precipitação de 1-12 meses após o nascimento (-2,88) é maior do que o efeito da precipitação de 13-24 meses antes do nascimento (-0,914), corroborando para as evidências da literatura de que os primeiros anos de vida dos indivíduos são períodos sensíveis²⁹ e respondem mais precisamente a choques adversos (Cunha e Heckman, 2007; Heckman, 2007). Por exemplo, privação nutricional no início da vida decorrentes das secas pode retrair indicadores antropométricos persistindo até a vida adulta (Aguilar e Vicarelli, 2011).

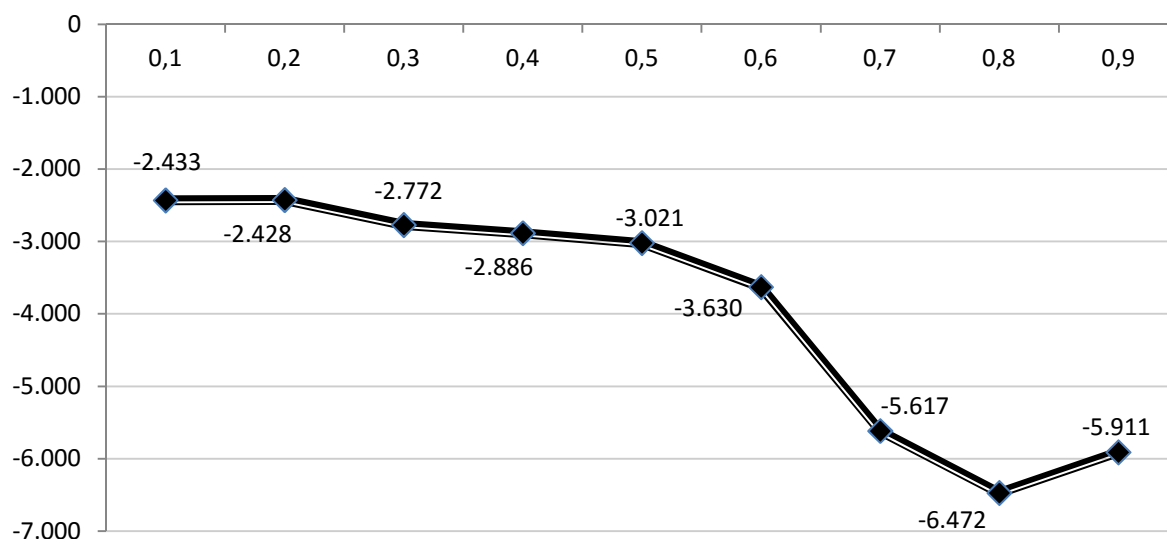
Quarto, e não menos importante, é notório que a falta de rigor metodológico imprime um papel demasiadamente relevante para o clima nos resultados de Rocha e Soares (2015). Notem que as colunas (1), (2) e (3) são, em números, quase o dobro das colunas (4), (5) e (6). Em síntese, tem-se um efeito superestimado da precipitação, o que nos leva a concluir *a priori* que existe condição endógena entre as variáveis relacionadas à precipitação que afetam resultados de saúde dos indivíduos, muito embora seja consensual de que esse efeito é negativo.

Sob outra ótica, observam-se resultados qualitativamente semelhantes para o Painel B que apresentam estimativas considerando uma *dummy* de seca. Por conseguinte, os resultados corroboram-se com os resultados do Painel A uma vez que mostra que as secas tem impacto positivo sob a taxa de mortalidade infantil dos quais tentam captar a gravidade dos choques destas. No geral nossos resultados confirmam os de Rocha e Soares (2015) de que independente de como se mensura a variável precipitação, são estatisticamente significativas e negativas. No entanto, considerando a heterogeneidade não observada verificou-se que as estimativas dos autores estavam superestimadas.

Outro ponto importante a destacar é que o método de Canay (2015) nos permite ir além das estimativas condicionais a média e conhecer um pouco da distribuição condicional através dos *quantis*. O Gráfico 3.1 apresenta as estimativas controladas pela temperatura e precipitação nos 12 meses antes 24 meses após do nascimento ao longo da distribuição, no qual ressalta-se que ambas apresentaram-se estatisticamente significativas.

²⁹ De acordo com Cunha e Heckman (2007) e Heckman (2007) mostram que etapas importantes no desenvolvimento físico dos indivíduos pode ocorrer na infância e ter consequências para os resultados de saúde quando adultos. São os chamados períodos sensíveis que segundo os autores decorrem da primeira infância que vai até os primeiro quatro anos de idade.

Gráfico 3.1 – Estimativas do impacto das secas sob os resultados dos indivíduos ao longo da distribuição



Fonte: Elaboração Própria com os resultados das estimações.

Nota: Todas as estimativas foram são estatisticamente significativos.

Verifica-se que o impacto da porcentagem da precipitação média cresce negativamente na medida em que cresce a taxa de mortalidade infantil. Em outras palavras, é possível perceber que o efeito é negativo, com tendência a aumentar na medida em que avançamos através dos *quantis*. É interessante ressaltar que sob o ponto de vista de políticas públicas, esta evidência tem papel importante ao identificar que choques externos tem efeito de longo prazo e com intensidade diferenciada (Maccini, 2009).

As evidências também implicam em dizer que além dos resultados diretos em saúde, poder-se-ia também alcançar efeitos indiretos em educação comprometendo as habilidades cognitivas dos indivíduos (Heckman, 2007). Embora as funções cognitivas possam ser afetadas negativamente por herança epigenética, a baixa cognição pode ocorrer, também, pela ausência de condições favoráveis durante a gravidez e na infância (Currie, 2006). Por exemplo, a escassez de água pode reduzir a produção agrícola e a ingestão suficiente de nutrientes afetando os resultados de saúde, e a falta de água potável pode gerar maior incidência de doenças infecciosas comprometendo a qualidade de vida da região (Rocha e Soares, 2015).

Com isso, políticas voltadas a atenuar os efeitos da seca na vida intrauterina tendem a melhorar o bem-estar dos indivíduos no longo prazo, e, quanto maior o *quantil* pertence o município, melhor seria o efeito desta política pública tanto em termos de saúde quando em educação no futuro dos indivíduos. Esta conclusão abre espaço para uma agenda de pesquisa em economia da saúde e da educação que foge ao escopo deste artigo.

3.3.2 Endogeneidade

Com o intuito de ampliar a análise esta subseção baseia-se no método proposto por Harding e Lamarche (2009) visando endogeneizar a quantidade de precipitações através de um modelo empírico de regressão quantílica de dados em painel com variáveis instrumentais, ora aqui chamado de estimador QR(IV). Os resultados encontram-se na Tabela 3. A inserção de variáveis instrumentais dado que estas não pertencem à equação explicativa, mas está correlacionada com a variável de precipitação é aqui justificada pelos fenômenos oceano-atmosfera ENOS (Rao, Lima e Franchito, 1993) e do Dipolo Atlântico (Aragão, 1998), expressos na variável Temperatura da Superfície do Mar (TSM).

Tabela 3.3 - Resultados das estimações com método de Harding e Lamarche (2009) do período de 1996-2010 da Região Semiárida do Nordeste

	QR(FE)			QR(VI)		
	(7)	(8)	(9)	(10)	(11)	(12)
Painel A – Precipitação						
Precipitação antes do nascimento	1,09*** (0,129)	-2,08*** (0,134)	-3,02*** (0,141)	0,65*** (0,134)	-2,49*** (0,132)	-2,63*** (0,133)
Precipitação 13-24 meses antes do nascimento			-0,91*** (0,127)			-1,94*** (0,126)
Precipitação 1-12 meses após o nascimento			-2,88*** (0,142)			-1,98*** (0,131)
Painel B - Seca						
Seca antes do nascimento	18,83*** (0,231)	2,67*** (0,147)	2,95*** (0,161)	18,81*** (0,2358)	2,83*** (0,152)	2,86*** (0,156)
Precipitação 13-24 meses antes do nascimento			-1,15*** (0,125)			-2,03*** (0,127)
Precipitação 1-12 meses após o nascimento			-2,69*** (0,132)			-1,79*** (0,120)
Observações	182.458	182.458	169.946	182.458	182.458	169.946
Número de Municípios	1.048	1.048	1.048	1.048	1.048	1.048

Fonte: Elaboração dos autores com base nas estimações.

Nota 1: Erros padrão robustos entre parênteses. ***p-valor < 0,01. ** p-valor < 0,05. * p-valor < 0,10.

Nota 2: Os resultados por *quantis* encontram-se no anexo I.

São transportados para a Tabela 3.3, colunas (7), (8) e (9) os resultados já apresentados nas colunas (4), (5) e (6) da Tabela 3.2 no que concernem as estimativas obtidas através de Canay (2011). Por sua vez, as colunas (10), (11) e (12) mostram os resultados para o estimador EQD(IV), que intuitivamente tem as mesmas conclusões do estimador QR(FE), exceto pelas magnitudes dos parâmetros que diminuem em absoluto. Enquanto que Rocha e Soares (2015) concluíram que aumentos de precipitação nos primeiros anos de vida pode reduzir em torno de 6% a taxa de mortalidade infantil, o estimador EQD(IV) demonstra que seu poder redutor resume-se em apenas 2,6%. Tais evidências são encontradas também por *quantil*, conforme Anexo I, no qual as variáveis instrumentais foram significativas.

Sob outro prisma, o Painel B apresenta uma variante do nível de precipitação. Denota-se a variável seca uma *dummy* que quando assumido valor 1 significa que a precipitação ao

longo dos 12 meses anteriores ao nascimento era mais do que um desvio padrão abaixo da média histórica. Com esta medida a interpretação se inverte, espera-se sinal positivo desta relação, conforme se verifica nos resultados dos estimadores aqui utilizados. Contudo, análogo à análise anterior, constata-se que os resultados de Rocha e Soares (2015) também foram superestimados provavelmente pelos problemas de endogeneidade não contabilizados.

Em suma, apesar dos resultados entre os estimadores serem semelhante qualitativamente, o que corrobora para a aplicação de políticas públicas voltadas a amenizar os efeitos das secas, é relevante destacar que as evidências correspondentes apontam um trato demasiadamente elevado ao *status* clima nos resultados de saúde dos indivíduos. Ademais, suspeita-se que esse *quórum* superestimado dos parâmetros possa, além do método inadequado, também ser decorrente do uso inapropriado de dados de precipitação, conforme analisado na próxima subseção.

3.3.3 Dados Observacionais

É consensual que dados históricos de chuvas e temperatura são boas medidas para verificar seu impacto sobre os resultados de saúde dos indivíduos, principalmente em regiões como a semiárida brasileira que é acometida por longos períodos de secas, e, que, consequentemente deve gerar ônus para sua população residente. Uma das justificativas para seu uso é que a variabilidade destas afetam diversas atividades econômicas (Freitas, Franchito e Rao, 2010), o que potencializa seu impacto, principalmente em regiões subdesenvolvidas e dependentes da agricultura (Cirilo, 2008).

Um dos desafios centrais da literatura ao analisar a relação entre a precipitação no início da vida e os próprios resultados de saúde dos indivíduos é que os dados de precipitação utilizados são suscetíveis a erro. Mas como isso ocorre? Isto porque o nível de precipitação é obtido através de estações meteorológicas que em virtude das dificuldades geográficas e do custo de mantê-las, muitas vezes não abrangem um contingente territorial representativo. Além disso, mesmo em locais onde existam estações meteorológicas, pelas mais diversas razões, pode-se gerar problemas que as tornem cronologicamente descontínua (Pinto et. al., 2009). Para contornar este problema, surgem na literatura às chamadas séries de reanálise baseados em modelos matemáticos, como por exemplo, interpolação linear, entre outros. No entanto, a contra gota é que essa medida também contém erros que podem superestimar as variabilidades climatológicas estudadas.

Os dados de reanálise são modelados com a média entre dois pontos de grade distanciados entre si em diferentes resoluções espaciais baseados nas distâncias de latitude e longitude (equivalente a km). Quanto maior a resolução, maior a distância entre os pontos. No entanto, as diferenças entre Observacionais e Reanálise poderiam decorrer das diferentes representatividades dos dados utilizados, ora aqui questionado. Para consolidar as conclusões aqui expostas, utilizou-se os dados de reanálise a *Terrestrial Air Temperature and Terrestrial Precipitation: 1900–2010 Gridded Monthly Time Series*, versões 3.01 e 3.02, respectivamente

(Matsuura and Willmott, 2009) utilizados em Rocha e Soares (2015), considerando apenas os municípios do semiárido em que tem informações na base de dados observacional.

No que tange a escolha dos dados observacionais, optou-se pelos dados obtidos no BDMEP, para testar se erro de medição clássica na variável precipitação no início da vida leva a estimativa dos coeficientes atenuados. Muito embora a escolha da base implique em redução da amostra estudada, não há embargo uma vez que estamos comparando mesmos municípios. Portanto, resume-se em 47 municípios analisados em uma amostra de 8.070 observações. No que se refere a questão metodológica, optou-se pelo método de Harding e Lamarche (2009) uma vez já comprovados nas seções anteriores como mais adequado para estimar a relação aqui investigada em virtude deste captar a heterogeneidade não observada e tratar a endogeneidade existente.

Tabela 3.4 - Resultados das estimações de comparação de dados de Reanálise e Observacionais do período de 1996-2010 da Região Semiárida do Nordeste

	Dados de Reanálise			Dados Observacionais		
	(13)	(14)	(15)	(16)	(17)	(18)
Painel A – Precipitação						
Precipitação antes do nascimento	-5,79*** (0,897)	-4,60*** (0,958)	-4,26*** (1,020)	-4,42*** (0,924)	4,56*** (0,972)	-3,80*** (1,040)
Precipitação 13-24 meses antes do nascimento			0,27 (0,956)			-2,83** (1,124)
Precipitação 1-12 meses após o nascimento			-2,52*** (0,944)			-7,72*** (1,239)
Painel B – Seca						
Seca antes do nascimento	20,44*** (1,361)	4,36*** (1,194)	5,52*** (1,683)	10,72*** (1,477)	4,35*** (1,454)	4,09*** (1,385)
Precipitação 13-24 meses antes do nascimento			0,34 (0,981)			-2,73** (1,126)
Precipitação 1-12 meses após o nascimento			-1,77* (0,950)			-7,52*** (1,178)
Observações	8.070	8.070	7.541	8.070	8.070	7.541
Número de Municípios	47	47	47	47	47	47

Fonte: Elaboração dos autores com base nas estimações.

Notas: Erros padrão robustos entre parênteses. ***p-valor < 0,01. ** p-valor < 0,05. * p-valor < 0,10. As colunas (13) e (16) não considera a temperatura antes do nascimento nas estimações.

A sinopse dos resultados é apresentada na Tabela 3.4, dos quais as colunas (13), (14) e (15) com estimativas dos dados de reanálise, enquanto que as colunas (16), (17) e (18) expõe os dados observacionais. Para ambos os resultados, considerou-se a Temperatura do Oceano Atlântico (TSM) como variável instrumental. Observa-se que a maioria dos resultados foram estatisticamente significativos e negativos, ratificando os resultados anteriores, contudo, merece atenção alguns fatos.

A primeira conclusão, *a priori*, é que de fato variáveis oriundas de reanálise tendem a superestimar os parâmetros desta relação. Notem que para cada coluna o parâmetro estimado com dados de reanálise são maiores do que aqueles estimados com dados de precipitação

observacionais. Com isso, vale aqui destacar duas importantes reflexões. Primeiro, chama-se atenção para a Tabela 3.4, coluna (15), que o controle precipitação de 13-24 meses antes do nascimento não é estatisticamente significativo nos dois estimadores (variável precipitação e seca), dando apenas importância para a precipitação após a concepção dos indivíduos. Parece que reduzir a amostra potencializa erro de medição da variável comprometendo os resultados.

A segunda reflexão se concentra na coluna (17), em que apresenta coeficiente positivo. Quando não controlado pelas variáveis de precipitação antes e depois do nascimento, o resultado é contraditório. Seria plausível pensar que como estamos em um número reduzido de municípios, talvez para os municípios pertencentes a amostra, esse resultado fosse amplamente aceito, no entanto, a coluna (18) mostra ser conivente com a literatura e com os demais resultados aqui explorados de que aumento de chuvas no período de nascimento tendem a reduzir a taxa de mortalidade infantil.

Com vistas a esses resultados, frisa-se que os efeitos de tratamento desse estudo ratifica o efeito negativo entre precipitação e mortalidade infantil, contudo, em magnitude inferior ao apresentado em Rocha e Soares (2015).

3.4 CONSIDERAÇÕES FINAIS

O objetivo do trabalho foi avaliar o impacto dos resultados pluviométricos e de períodos de seca na taxa de mortalidade infantil no semiárido nordestino, a fim de ratificar ou não os resultados encontrados na literatura. Para tanto, utilizou-se três abordagens: *i)* método de regressão quantílica com dados em painel de efeitos fixos relatado em Canay (2011); *ii)* método de regressão quantílica com dados em painel com variáveis instrumentais proporcionado em Harding e Lamarche (2009) utilizando a Temperatura do Oceano como variável instrumental; *iii)* dados observacionais para contrapor os de reanálise.

Os resultados indicam um efeito negativo da precipitação e um efeito positivo da seca na taxa de mortalidade infantil, para as três estratégias. No entanto, a ausência de tratamento no que concerne considerar a heterogeneidade não observada e a endogeneidade, bem como o uso de dados modelados, revelam possíveis superestimações. Dessa forma, os efeitos de tais fenômenos naturais na redução da mortalidade infantil existem, mas o seu impacto pode ser menor do que o comumente aceito, sendo a maior parte do impacto devido a outros fatores, tais como as condições de saúde dos municípios, entre outros.

Logo, as principais conclusões são: *i)* quanto maior a precipitação antes da concepção, durante o período gestacional e nos primeiros anos de vida dos indivíduos nascidos no semiárido menor a taxa de mortalidade infantil nesta região; *ii)* quando controlado a heterogeneidade não observada e a endogeneidade existente nesta relação, a precipitação tem menor peso indicando que existem outras variáveis que afetam os resultados de saúde; *iii)* o método utilizado em Rocha e Soares (2015) não é adequado para estimar essa relação. Sem embargo, o Painel B, que considera a variável seca, apresentam as mesmas conclusões sob ângulo oposto.

Em síntese, temos que bebês nascidos nas zonas áridas que sofrem secas quando no útero de suas mães residentes no semiárido enfrentam risco maior de morte, especialmente se nasceram quando as secas comprometem os seus primeiros anos de vida e quando a escassez de água afeta a nutrição materna antes da concepção desses indivíduos. Trabalhos como o de Maccini e Yang (2009), Kudamtsu et. al (2010); Aguilar e Vicarelli(2011), Shah e Steberg (2013), Burgess et. Al (2014) e Rocha e Soares (2015) tem ressaltado os efeitos das questões climáticas nos resultados de saúde, contudo, este estudo vem a contribuir ao ressaltar a importância de escolher métodos adequados que tratem os problemas existentes, bem como a escolha de dados confiáveis e que sejam fidedignos da relação causal a ser estudada.

3.5 REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, José M. Brabo; SOUZA, R. O.; CAMPOS, José Nilson B. Previsão da anomalia de temperatura da superfície do mar (tsm) no atlântico tropical, com a equação da difusão de temperatura. **Revista Climanalise**, ano, v. 3, p. 163-172, 2003.

ANDREOLI, Rita Valéria; KAYANO, Mary Toshie. A importância relativa do atlântico tropical sul e pacífico leste na variabilidade de precipitação do Nordeste do Brasil. **Revista Brasileira de Meteorologia**, v. 22, n. 1, p. 63-74, 2007.

AGUILAR, Arturo; VICARELLI, Marta. El Nino and Mexican children: medium-term effects of early-life weather shocks on cognitive and health outcomes. **Cambridge, United States: Harvard University, Department of Economics. Manuscript**, 2011.

ARAGÃO, José Oribe Rocha. O Impacto do ENSO e do Dipolo do Atlântico no Nordeste do Brasil. **Bull. Inst. fr. études andines**, v. 27, n. 3, p. 839-844, 1998.

ARAÚJO COSTA, José. O FENÔMENO EL NIÑO E AS SECAS NO NORDESTE DO BRASIL. **EDUCTE: Revista Científica do Instituto Federal de Alagoas**, v. 3, n. 1, 2014.

BARKER, D. J. P. **Mothers, Babies and Health in Later Life**. Editora Churchill Livingstone, edição nº 2, ISBN-10: 0443061653, 1998.

BRAHMANANDA RAO, V.; DE LIMA, Marley C.; FRANCHITO, S. H. Seasonal and interannual variations of rainfall over eastern northeast Brazil. **Journal of Climate**, v. 6, n. 9, p. 1754-1763, 1993.

BEZABIH, Mintewab; DI FALCO, Salvatore; MEKONNEN, Alemu. **On the Impact of Weather Variability and Climate Change on Agriculture: Evidence from Ethiopia**. 2014.

BURGESS, Robin et al. The unequal effects of weather and climate change: Evidence from mortality in india. **Unpublished working paper**, 2013.

CANAY, Ivan A. A simple approach to quantile regression for panel data. **The Econometrics Journal**, v. 14, n. 3, p. 368-386, 2011.

CIRILO, José Almir. Public water resources policy for the semi-arid region. **estudos avançados**, v. 22, n. 63, p. 61-82, 2008.

CUNHA, Flavio; HECKMAN, James. **The technology of skill formation**. National Bureau of Economic Research, 2007.

CURRIE, Janet. **Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development**. National Bureau of Economic Research, 2008.

CURRIE, Janet et al. Something in the water: Contaminated drinking water and infant health. **Canadian Journal of Economics/Revue canadienne d'économique**, v. 46, n. 3, p. 791-810, 2013.

CURRIE, Janet; VOGL, Tom. **Early-life health and adult circumstance in developing countries**. National Bureau of Economic Research, 2012.

CURRIE, Janet; ROSSIN-SLATER, Maya. Weathering the storm: Hurricanes and birth outcomes. **Journal of health economics**, v. 32, n. 3, p. 487-503, 2013.

FIORI, N. As neurociências cognitivas. Tradução de Sonia M. S. Fuhrmann. Petrópolis RJ:Vozes, 2008.

FREITAS, Ana Carolina Vasques; FRANCHITO, Sérgio Henrique; RAO, Vadlamudi Brahmananda. Analise dos dados de precipitacao provenientes de diferentes fontes sobre a américa do sul, com enfase no brasil. **CLIMEP-Climatologia e Estudos da Paisagem, Rios Claros, PR, Brasil**, v. 5, n. 1, 2010.

GALVAO, Antonio F.; LAMARCHE, Carlos; LIMA, Luiz Renato. Estimation of censored quantile regression for panel data with fixed effects. **Journal of the American Statistical Association**, v. 108, n. 503, p. 1075-1089, 2013.

GLEWWE, Paul; KING, Elizabeth M. The impact of early childhood nutritional status on cognitive development: Does the timing of malnutrition matter?. **The world bank economic review**, v. 15, n. 1, p. 81-113, 2001.

GLUCKMAN, Peter. e HANSON, Mark. **The fetal Matrix: Evolution, Development, and Disease**. Cambridge: Cambridge University Press, 2005.

GROSSMAN, Michael. The demand for health: a theoretical and empirical investigation. **NBER Books**, 1972.

GUEDES, R et al. Série temporal de precipitação mensal de Fortaleza, Brasil: Comparação entre observações e dados de reanálise do NCEP/NCAR. **Revista Brasileira de Meterologia**, v. 20, p. 83-92, 2005.

HARDING, Matthew; LAMARCHE, Carlos. A quantile regression approach for estimating panel data models using instrumental variables. **Economics Letters**, v. 104, n. 3, p. 133-135, 2009.

HASTENRATH, S.; HELLER, L. Dynamics of climatic hazards in north-east Brazil. *Quartely Journal of the Royal Meteorological Society*, v. 110, p. 411-425, 1977.

HECKMAN, James. The technology and neuroscience of skill formation. **Invest in Kids Working Group, Center for Economic Development, Partnership for America's Economic Success, July**, v. 17, 2007.

KAYANO, Mary Toshie; RAO, Vadlamudi Brahmananda; MOURA, Antonio Divino. Tropical circulations and the associated rainfall anomalies during two contrasting years. **Journal of Climatology**, v. 8, n. 5, p. 477-488, 1988.

KOENKER, Roger. Quantile regression for longitudinal data. **Journal of Multivariate Analysis**, v. 91, n. 1, p. 74-89, 2004.

KUDAMATSU, Masayuki; PERSSON, Torsten; STRÖMBERG, David. Weather and Infant Mortality in Africa. 2012. Manuscrito não publicado.

MACCINI, Sharon L.; YANG, Dean. **Under the weather: Health, schooling, and economic consequences of early-life rainfall**. National Bureau of Economic Research, 2008.

MORAES NETO, João M.; BARBOSA, Marx P.; DE ARAÚJO, Alexandre E. Efeito dos eventos ENOS e das TSM na variação pluviométrica do semi-árido paraibano. **R. Bras. Eng. Agríc. Ambiental**, v. 11, n. 1, p. 61-66, 2007.

MOURA, A. D.; SHUKLA, J. On the dynamics of the droughts in Northeast Brazil: observations, theory and numerical experiments with a general circulation model. *Journal Atmospheric Science*, v. 38, n. 12, p. 2653-2673, 1981.

PINTO, L. et al. Comparação de produtos de reanálise para a América do sul. **Revista Brasileira de Meteorologia**, v. 24, n. 4, p. 461-472, 2009.

ROCHA, Rudi; SOARES, Rodrigo R. Water scarcity and birth outcomes in the Brazilian semiarid. **Journal of Development Economics**, v. 112, p. 72-91, 2015.

SANTOS, C.; BRITO, J. Análise dos índices de extremos para o semi-árido do Brasil e suas relações com TSM e IVDN. **Revista Brasileira de Meteorologia**, v. 22, n. 3, p. 303-312, 2007.

SHAH, Manisha; STEINBERG, Bryce Millett. Could droughts improve human capital? evidence from india. **Unpublished manuscript, last modified**, 2012.

SILVA, Viviane BS et al. An improved gridded historical daily precipitation analysis for Brazil. **Journal of Hydrometeorology**, v. 8, n. 4, p. 847-861, 2007.

TRENBERTH, K. E. The definition of El Niño. **Bulletin of the American Meteorological Society**, v.78, n.12, p.2771-2777. 1997.

UGAZ, Jorge; ZANOLINI, Arianna. Effects of extreme weather shocks during pregnancy and early life on later health outcomes: the case of philippines? typhoons. 2011.

XAVIER, T. M. B. S.; XAVIER, A. F. S.; SILVA-DIAS, M. A. F.; SILVA-DIAS, P. L. Interrelações entre Eventos ENOS (ENSO), a ZCIT (ITCZ) no Atlântico e a Chuva nas Bacias Hidrográficas do Ceará, **Revista Brasileira Recursos Hídricos**, v. 8, n. 2, p. 111-126. 2003.

Anexo I – Resultados por *quantil* do Estimador QRD(FE)

<i>Quantil</i>	<i>Coefficiente</i>	<i>IV atlântico</i>	<i>IV pacífico</i>
0.1	-1,521*** (0,134)	-0,315*** (0,033)	0,212*** (0,028)
0.2	-1,729*** (0,124)	-0,316*** (0,030)	0,210*** (0,026)
0.3	-2,279*** (0,108)	-0,355*** (0,029)	0,214*** (0,024)
0.4	-2,405*** (0,110)	-0,368*** (0,029)	0,213*** (0,023)
0.5	-2,627*** (0,133)	-0,414*** (0,033)	0,220*** (0,026)
0.6	-3,457*** (0,172)	-0,455*** (0,042)	0,180*** (0,034)
0.7	-6,310*** (0,311)	-0,648*** (0,082)	0,075 (0,071)
0.8	-8,143*** (0,736)	-1,604*** (0,219)	-0,486*** (0,166)
0.9	-4,965*** (1,531)	-3,186*** (0,487)	-1,907*** (0,352)

Fonte: Elaboração dos autores.

CONSIDERAÇÕES FINAIS

Na área de economia, sobretudo em econometria aplicada, tem-se utilizado modelo de dados em painel. Contudo, a maioria dos estimadores são projetados para modelos com distribuição gaussiana e, apesar de captar a heterogeneidade não observada através dos efeitos fixos, não conseguem estimar os efeitos heterogêneos das covariáveis que são muitas das vezes importantes para o desenho das políticas públicas. Modelos de Regressão Quantílica, por sua vez, permitem conhecer a distribuição condicional dos efeitos, porém, os estimadores comumente utilizados não permitem a inclusão dos efeitos fixos para controlar covariáveis latentes. Neste sentido, a estratégia do presente trabalho é aplicar técnicas de Regressão Quantílica com estrutura de dados em painel que permite a inclusão dos efeitos fixos para três temas em economia social. Portanto, a contribuição da tese é adicionar elementos a literatura da educação e saúde com estimadores mais robustos.

Os resultados gerais sugerem um efeito superestimado das variáveis de interesse quando não contabilizados a heterogeneidade não observada e os efeitos das covariáveis. Neste contexto, encontrou-se no primeiro ensaio os seguintes resultados. O tamanho da turma tem um efeito negativo, mas não persistente entre os *quantis* de desempenho acadêmico. Tal constatação indica que políticas de redução de turma causariam efeitos benéficos em alunos com menores notas, pois, parte-se do pressuposto de que estes poderiam se beneficiar de maiores interações entre aluno e professor. Além disso, classes menores possibilitariam os docentes a se dedicarem mais tempo na preparação de materiais complementares as aulas expositivas e atendimento fora de sala de aula. Tal apontamento é ratificado pela covariável carga horária exercida no semestre pelo docente, onde apresenta um efeito negativo sobre o desempenho para todos os *quantis*, principalmente na cauda inferior. Por outro lado, considerando alunos de alta performance normalmente possuem maiores habilidades potencializadas pela interação com seus pares, e, portanto, redução do tamanho da turma os prejudicaria.

Já os efeitos das chuvas no período gestacional sobre os resultados de educação e saúde, foram investigados no segundo e terceiro ensaio, respectivamente. No que tange à educação, utilizou-se a base de dados da UFPB dos alunos que nasceram no semiárido com o objetivo de averiguar os efeitos das condições ambientais iniciais, notadamente o nível de precipitação média ocorrida no período intrauterino, sobre o desempenho acadêmico na faculdade. Os resultados mostraram uma relação positiva e persistente ao longo dos *quantis*. Ao considerar a variável dependente a taxa de mortalidade infantil como uma *proxy* para os resultados de saúde em um painel *município-a-mês*, o terceiro ensaio concluiu que quanto maior a precipitação antes da concepção, durante o período gestacional e nos primeiros anos de vida, menor são as taxas de mortalidade infantil dos municípios e que esse efeito tende a aumentar a medida em que avançamos através dos *quantis*, indicando que choques externos na vida intrauterina tem efeito no longo prazo e com diferentes intensidades.