

GOVERNO DO ESTADO DO CEARÁ
SECRETARIA DO PLANEJAMENTO E GESTÃO – SEPLAG
INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ – IPECE

ECONOMIA DO CEARÁ EM DEBATE 2020

ipece INSTITUTO
DE PESQUISA
E ESTRATÉGIA
ECONÔMICA
DO CEARÁ



CEARÁ
GOVERNO DO ESTADO
SECRETARIA DO
PLANEJAMENTO E GESTÃO

GOVERNO DO ESTADO DO CEARÁ
SECRETARIA DO PLANEJAMENTO E GESTÃO – SEPLAG
INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ – IPECE

ECONOMIA DO CEARÁ EM DEBATE 2020

Fortaleza - CE
2021

Governador do Estado do Ceará

Camilo Sobreira de Santana

Vice-Governadora do Estado do Ceará

Maria Izolda Cella de Arruda Coelho

Secretário do Planejamento e Gestão – SEPLAG

Flávio Ataliba Flexa Daltro Barreto – Secretário (respondendo)

Secretário Executivo de Planejamento e Orçamento – SEPLAG

Flávio Ataliba Flexa Daltro Barreto

Secretário Executivo de Gestão

Adriano Sarquis Bezerra de Menezes

Secretário Executivo de Planejamento e Gestão Interna

Ronaldo Lima Moreira Borges

Instituto de Pesquisa e Estratégia Econômica do Ceará – IPECE

Diretor Geral

João Mário Santos de França

Diretoria de Estudos Econômicos – DIEC

Ricardo Antônio de Castro Pereira

Diretoria de Estudos Sociais – DISOC

Diretoria de Estudos de Gestão Pública – DIGEP

Marília Rodrigues Firmiano

Gerência de Estatística, Geografia e Informação – GEGIN

Rafaela Martins Leite Monteiro

ECONOMIA DO CEARÁ EM DEBATE 2019

2020 v-1 - 2015

João Mário de França, Rafaela Martins Leite Monteiro, Fátima Juvenal de Sousa (organizadores).

Anual

ISBN: 978-65-990380-6-8

1. Economia. 2. Ceará. I - França, João, Mário.

CDU 330 (813.1) (05)

250 páginas. Copyright © 2021 - IPECE.

Os artigos apresentados neste livro são de inteira responsabilidade dos seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista do Instituto de Pesquisa e Estratégia Econômica do Ceará - IPECE.

SUMÁRIO

INTRODUÇÃO.....5

EFFICIENCY IN TAX COLLECTION: OVERCOMING THE EFFECTS OF THE COVID-19 PANDEMIC ON PUBLIC FINANCES OF THE CEARÁ STATE.....7

Francisco Germano Carvalho Lucio

Jhorland Ayala Garcia

Ricardo A. de Castro Pereira

MODELO DE EQUILÍBRIO GERAL COMPUTÁVEL E ANÁLISE INSUMO-PRODUTO: CONTRASTANDO APLICAÇÕES PARA O CEARÁ.....27

Christiano Modesto Penna

Ricardo de Castro Pereira

Nicolino Trompieri Neto

Witalo de Lima Paiva

O EFEITO DA PANDEMIA DE COVID-19 NOS GASTOS PÚBLICOS MUNICIPAIS NO ESTADO DO CEARÁ: PROPOSIÇÃO E APLICAÇÃO DE MÉTODO DE DETECÇÃO DE RED FLAGS.....44

Sarah Mesquita Lima

Jaime de Jesus Filho

Paulo Rogério Faustino Matos

Ricardo Wagner Cavalcante Brito

A QUALIDADE E EXPANSÃO DO ENSINO SUPERIOR CONTRIBUI PARA O *OVEREDUCATION*? EVIDÊNCIAS PARA O CEARÁ.....64

Celina Oliveira

Francisca Zilania Mariano

Jair Andrade Araújo

AN UNINTENDED CONSEQUENCE OF VOCATIONAL SCHOOLS' CONSTRUCTION: EVIDENCE FROM CEARÁ.....83

Rafael B. Barbosa

Alesandra Benevides

Francisca Zilânia Mariano

ANÁLISE DA INCIDÊNCIA E DO PERÍODO DE AFASTAMENTO POR ACIDENTES DE TRABALHO OU DOENÇAS OCUPACIONAIS NO ESTADO DO CEARÁ.....103

Mateus Mota dos Santos

Isabela Braga Sales

David Herbster Ferraz

ANÁLISE DE CONVERGÊNCIA ESPACIAL DE PRODUTIVIDADE AGRÍCOLA APLICADA À REGIÃO NORDESTE DO BRASIL.....	127
Winnie Moreira Albuquerque	
Roberio Telmo Campos	
Vitor Hugo Miro Couto Silva	
DESMATAMENTO DE MATA ATLÂNTICA NO ESTADO DO CEARÁ: ANÁLISES DA CURVA DE KUZNETS AMBIENTAL A PARTIR DE DADOS EM PAINEL, 2011 A 2017.....	148
Wesley Leitão de Sousa	
Guilherme Irffi	
Moisés Dias Gomes de Asevedo	
ESCOLARIDADE DOS PAIS E PERFORMANCE ESCOLAR DOS FILHOS AO FINAL DO ENSINO MÉDIO NO CEARÁ.....	168
Janaína Rodrigues Feijó	
Valdemar Rodrigues de Pinho Neto	
ESCOLAS EM TEMPO INTEGRAL E DESEMPENHO NO ENEM: UMA AVALIAÇÃO DE IMPACTO PARA O ESTADO DO CEARÁ.....	190
Edward Costa Martins	
Gerrio dos Santos Barbosa	
Vitor Hugo Miro Couto Silva	
Helson Gomes de Souza	
Luciane dos Santos Jácome	
IMPACTO DESIGUAL DO FINANCIAMENTO DA EDUCAÇÃO BASEADO EM RESULTADOS.....	208
Pedro Alexandre Santos Veloso	
Rafael Barros Barbosa	
QUEM SÃO OS JOVENS ‘NEM NEM’ NA REGIÃO METROPOLITANA DE FORTALEZA.....	229
Antonia Jaine da Silva Pereira	
Silvana Nunes de Queiroz	

INTRODUÇÃO

O livro *Economia do Ceará em Debate 2020* é uma coletânea dos doze artigos apresentados por ocasião do **XVI Encontro Economia do Ceará em Debate**, realizado pelo **Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE)**, no dia 04 de dezembro de 2020, em evento online com transmissão pelo YouTube.

O livro que apresentamos à sociedade é fruto do Encontro, bem como do trabalho interativo entre Governo do Estado e à academia, que contou com o envolvimento intenso dos colaboradores do **IPECE**. De forma a cobrir os principais temas de pesquisa em economia com foco no Estado do Ceará, e conceber oportunidade para que estudiosos possam discutir os principais desafios da economia cearense, bem como possibilitar às autoridades públicas responsáveis pelo desenvolvimento do Estado, conhecer ideias inovadoras sobre as ações que têm efeitos diretos sobre a economia e a sociedade cearense.

Agradecemos o apoio recebido do Governo do Estado do Ceará, por intermédio da **Secretaria do Planejamento e Gestão (SEPLAG)**.

João Mário Santos de França
Diretor Geral do IPECE

Organizadores

João Mário de França

Rafaela Martins Leite Monteiro

Fátima Juvenal de Sousa

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ (IPECE)

Av. Gal. Afonso Albuquerque Lima, s/n | Edifício SEPLAG | Térreo - Cambéa | Cep: 60.822-325

Fortaleza, Ceará, Brasil | Telefone: (85) 3101-3521

<http://www.ipece.ce.gov.br>

EFFICIENCY IN TAX COLLECTION: OVERCOMING THE EFFECTS OF THE COVID-19 PANDEMIC ON PUBLIC FINANCES OF THE CEARÁ STATE

Francisco Germano Carvalho Lucio*

Jhorland Ayala Garcia**

Ricardo A. de Castro Pereira***

ABSTRACT

The main contribution of this paper is to analyze the efficiency in tax collection as an alternative to overcome fiscal problems left by the Covid-19 pandemic, focusing on the Ceará state. A double analysis is performed. Firstly, it estimates the efficiency in the collection of ICMS and the total revenue for the Brazilian states. Then, it adapts a dynamic CGE model to consider these efficiencies, which is calibrated to represent the economy of the Ceará state. Based on the efficiency estimated, exercises were performed simulating improvements in efficiency levels in two strands, one increasing the efficiency in ICMS collection and the other combining it with an ICMS tax rate reduction. The latter is presented as an empirical demonstration for the axiom of the Compensation of the Tax Collection Drivers, proposed in this paper. Results show that the Ceará state presents one of the highest levels of efficiency in tax collection for both measures. Additionally, an improvement in tax collection efficiency would yield an increase in tax collection, continuously but diminishing over the years. We hope this work contributes to the public debate regarding efficiency in tax collection as a feasible alternative public policy to overcome the fiscal problems.

Keywords: Efficiency. Order- m Estimation. Tax Collection. Regional CGE Model.

JEL Codes: C68, H75, R50

RESUMO

A principal contribuição deste artigo é analisar a eficiência na arrecadação de impostos como alternativa para contornar os problemas orçamentários deixados pela pandemia da Covid-19, focando no Estado do Ceará. Para tanto, realizou-se uma análise dupla. Primeiramente, estima-se a eficiência na arrecadação do ICMS e da receita total dos estados brasileiros. Em seguida, adapta-se um modelo CGE dinâmico considerando esse tipo de eficiência, calibrado para representar a economia do estado do Ceará. Com base na eficiência estimada, realizaram-se simulações de melhorias nos níveis de eficiência em duas vertentes, uma aumentando a eficiência na arrecadação do ICMS e outra combinando-a com a redução da alíquota do ICMS. Esta apresenta-se como uma demonstração empírica do axioma da Compensação dos Determinantes da Arrecadação de Impostos, proposto neste artigo. Os resultados mostram que o estado do Ceará apresenta um dos maiores níveis de eficiência na arrecadação para ambas as medidas. Além disso, uma melhoria na eficiência da arrecadação produziria um aumento na arrecadação de impostos de forma contínua, mas decrescente ao longo do tempo. Espera-se que este trabalho contribua para o debate acerca da eficiência na arrecadação de impostos como alternativa viável de política pública para contornar os problemas fiscais.

* Department of Economics, CAEN – UFC, Ceará – Brazil. E-mail: germanocarvalho@caen.ufc.br. Tel. (85) 99996-6285.

** Agricultural and Consumer Economics – ACE Department, UIUC, IL – US. E-mail: ja8@illinois.edu.

*** Department of Economics, CAEN – UFC, Ceará – Brazil. and IPECE. E-mail: rpereira@caen.ufc.br.

Palavras-Chave: Eficiência. Estimação Order- m . Tributação. Modelo CGE Regional.

Classificação JEL: C68, H75, R50.

Acknowledgements

The authors thank the Ceará's Institute for Research and Economic Strategy – IPECE for the provision of a database, crucial for the development of this paper, and the Brazilian Federal Agency for Support and Evaluation of Graduate Education – CAPES.

1 INTRODUCTION

The COVID-19 pandemic that has plagued worldwide since the beginning of the year 2020 has led several governments both national and local, to declare a situation of emergency, consequently adopting measures that promote social distancing. Therefore, for a few months beginning in March, the main measures imposed restrictions on economic activities, given the fact that not only the workplaces but also the common daily activities usually gather people.

Given the fact that this health crisis has shown impacts in the whole world, a considerable amount of studies has been developed and published so far is already reported in the literature in several different fields. Specifically for economics, for instance, there are works for the economic effects from it (McKibbin and Fernando (2020), Fernandes (2020), and Atkeson (2020)) for financial markets (Zhang et al. (2020) and Ashraf (2020)), for poverty (Sumner et al. (2020), Buheji et al. (2020), and Martin et al. (2020)). For the Brazilian case specifically, these impacts are reported for instance by Domingues, Cardoso, and Magalhães (2020) and Domingues et al. (2020).

Regarding the economic effects in the Ceará state specifically, IPECE (2020a), using an indicator for the trend of the performance in the short-term, the Quarterly GDP, shows that the GDP of the Ceará state in the second quarter of 2020, the period with more severe measures to reduce the speed the virus spreads out, has decreased 14.55% comparing to the same period of 2019. Meanwhile, Brazil presented an 11.4% decrease. Considering the cumulative first semester, this decrease for Ceará represented 7.58%, while Brazil presented 5.9%. However, considering the cumulative from the last four quarters, the decrease is around 2.72% and 2.2% for Ceará and Brazil, respectively. So far the forecast for the economic growth of Ceará's economy for the year 2020 is -4.5%. It is better than the expectation for the Brazilian economy as a whole. According to IFI (2020), the government itself maintained its projection of a 4.7% decline and the International Monetary Fund (IMF) updated its projection to a retraction by 5.8%. This value is also close to the market expectations estimated by the Central Bank of a 5% drop. Finally, the IFI's fiscal scenarios consider a 6.5% decline, with an improvement bias.

Specifically in terms of the job market, based on the data published by PNAD COVID from IBGE for the period from May to August, IPECE (2020c) shows the estimated level of occupation in Ceará, which is the share of the population aged 14 and over who worked even in some informal occupation, fell from 41.6% in May to 39.9% in August. The estimated participation rate, which is the percentage of people in the workforce in the reference week in relation to people of working age, in turn, went from 45.6% in May to 46.0% in August. The unemployment rate, which is the percentage of unemployed people in the reference week in relation to people in the workforce that week, increased from 8.6% in August to 13.1%.

Revenues fall due to the retraction of economic activity and tax exemptions and deferrals. At the same time, expenditures are rising rapidly to mitigate the effects on the income of the most vulnerable part of the population and the financial situation of smaller companies, besides the increase in spends directly on the health system (IFI, 2020). According to official

data from IPECE (2020b), in April 2020 it was observed a significant reduction in the collection of ICMS (Tax on the Circulation of Goods and Services), which is the main source of revenue for the Ceará state government and accounts for approximately half of the current tax collection. It shows it was 36.7% lower than in the same month of the previous year. The biggest drop was in May 2020, with a reduction of 39.4%. Subsequently, in June and July there was a decrease in revenue, but with a lower intensity than in previous months. Considering the four months, which is the period that lasts the policy of social distancing and that it starts to be relaxed, the reduction in the collection of ICMS was approximately 22.7%.

Based on that, the Emergency Aid, implemented by the Complementary Law 173/2020¹ and provided by the Central Government, as well as extraordinary credits for the Ministry of Health are very close to this reduction in tax collection. Indeed, according to IPECE (2020d), the amounts transferred in order to mitigate the effects of Covid-19 outweigh the observed drop in the collection of ICMS. However, it should be noted that the second main source of state revenue, the State Participation Fund (FPE), consists of revenues from the Income Tax and taxes on industrialized products collected by the Federal government and transferred to the subnational units, and that are sensitive to the economic cycle. Indeed, the transfers carried out were lower than the revenue losses for the period in 163 million. In terms of net impacts, it can be concluded that the policies taken by the federal branches, Executive and Legislative, during the health crisis contributed significantly to mitigate their negative fiscal effects.

Even though the compensation from the Central Government has been enough for a while, one must remember there is still an expected reduction in the GDP for the current year of 2020. Hence, there will be reduction in tax collection. Moreover, it should be noted that there is a possibility that the tax collection will not recover quickly after the end of the pandemic period. It means that in the coming months the tax collection at all, not only the ICMS collection, will be at levels lower compared to the previous year.

The very first way that comes up in one's mind to surpass this upcoming and delicate fiscal situation is increasing tax rates. However, in terms of the tax burden, Brazil presents a share of the GDP similar to the average of OECD's members. Official data shows it was around 33.5% of the GDP in 2018, considered as one of the highest tax burdens in the world. Additionally, this amount collected has been not enough to afford public spends and it is facing successive fiscal deficits since 2014, contributing to a continuous enlargement of the public debt. Additionally, according to Romer and Romer (2010), tax increases are highly contractionary. Therefore, it is highly likely that it would worsen the current economic situation hence it is unlikely to be approved in congress.

Based on Brazilian federalism, the way the central government collects and distributes these revenues to the subnational units, this scenario combined with the remaining effects of the Covid-19 pandemic might bring future fiscal problems for the public finances of state-level governments as well. Therefore, an increase in efficiency in tax collection might provide an alternative way to mitigate the remaining effects of the Covid-19 pandemic in public finances.

Based on the described scenario, this paper proposes a potential alternative to solve or at least mitigate this problem. The alternative is increasing the efficiency of tax collection. The literature of efficiency performance in tax collection issues has not been properly explored, even though very recent literature has analyzed the efficiency on taxation, especially looking

¹ It defines the Federative Program for Confronting the Coronavirus SARS-CoV-2 (Covid-19), which basically suspends the payments of contracted debts between the central government and subnational units, transfers from National funds as financial assistance to the States, the Federal District, and the Municipalities, in the year 2020, to apply in actions to confront the SARS- Coronavirus CoV-2 (Covid-19).

for explanatory factors (Morrissey et al. (2016), Aizenman and Jinjark (2018), Cevik et al. (2019)). However, there is a gap in the literature combining efficiency on tax collection and dynamic CGE models, which allow the analysis to cover the whole economy. Therefore, this paper proposes to estimate the efficiency in tax collection for the Brazilian states, focusing on the Ceará state, and apply it in a regional and dynamic CGE model to simulate these sort of policies in the economy of Ceará. Additionally, it also simulates a compensation between an increase in efficiency and a reduction in tax rates.

Concerning the national literature about regional CGE, there are several references. Based on ORANI type models (Dixon et al., 1978; 1982), Guilhoto (1995) developed a model for Planning and Analysis for Agricultural Policies, which works for the State of São Paulo as well as for Brazil as a whole. Yet, based on Dixon e Parmenter (1996), Haddad (1999) created the first inter-regional CGE model totally operational for Brazil, called B-MARIA (Brazilian Multisectorial And Regional/Inter-Regional Analysis). From it, several other models have been developed such as the B-MARIA-SP (Domingues and Haddad (2002)), the B-MARIA-27 (Haddad (2004)), the B-MARIA-27-IT (Haddad and Perobelli (2005)) and the B-MARIA-RS (Porsse (2005)).

Domingues e Haddad (2002) and Domingues e Lemos (2004) splits Brazil into two regions, namely that, the state of São Paulo and the rest of Brazil. This is way easier than models based on TERM, from Horridge, Madden e Wittwer. (2003), which disaggregates for more regions. Aiming to evaluate the long-term effects of the installation of petroleum refineries in the Northeast region, Ribeiro (2017) creates a dynamic model called the Brazilian Northeast Inter-regional Model - B-NORIM. Fochezatto (2002) and, more recently, Braatz et al. (2015) for the Rio Grande do Sul state, which is both dynamics.

It is emphasized that none of the models presented in these references considers efficiency in the public sector, except for Lucio et al. (2019), however, it refers to a static model. Therefore, this paper aims to contribute to the literature by filling this gap, which means, by creating a dynamic and regional CGE model that considers the efficiency of the public sector in terms of tax collection. Additionally, once the governments influence the other economic agents by incentives, one way the government might take action is by increasing its efficiency level performing its duties, which certainly includes tax collection. Therefore, this approach seems suitable.

This paper is composed of five more sections, beyond this introduction. The following two sections present, respectively, the dynamic CGE model for the Ceará state and the calibration of it. In section four, the efficiency scores are estimated and minor analyses are performed as partial results. In section five, some simulations of changes in efficiency in tax collection are performed and discussed. Finally, the most relevant results are highlighted as well as some policy implications are provided.

2 THE ITERATIVE, RECURSIVE-DYNAMIC, AND CEARÁ'S ECONOMY MODEL OF ANALYSIS – IRACEMA.

The model developed in this paper is based on a recursive dynamic CGE model, from Hosoe et al. (2016). The Iterative, Recursive-dynamic, and Ceará's Economy Model of Analysis – IRACEMA² brings forth two major changes compared to its reference model. Firstly, once it was developed for state-level analysis, the foreign sector is divided into two regions, namely that, the rest of Brazil and the rest of the world, as in a static version presented in Lucio et al. (2019), which in turn is based on Hosoe et al. (2010). Although this characteristic is unnecessary for the intended approach, it yields a more realistic model for the type of region is being treated.

Besides, this model can be used as a reference model that considers differences in foreign trade, in which one region is subject to the exchange rate but the other is within a national trade, which in turn can be either a state-level approach or a region using a similar currency such as the European Union. Secondly, for covering the main point of the paper, efficiency parameters are included in the taxation duties of the government. The improvement proposed stands as a contribution not only to the national literature but also to the CGE literature as a whole. The IRACEMA Model is presented in detail in the following sections.

2.1 Productive Sectors

As long as the Iracema Model is dynamics, this feature must be emphasized in this beginning. Thus, the factor endowments, labor, and capital evolve over the periods. The labor endowments grow at the same rate as the population growth (pop), then it is driven by this equation:

$$FF_{LAB,t+1} = (1 + pop) * FF_{LAB,t} \quad (1)$$

Once installed the capital is assumed immobile among the sectors. Let the parameter dep represent the depreciation of capital, the capital stock in sector j in the following period is determined by the combination of the current capital stock combined with new capital, converted from investment ($II_{j,t}$). It can be represented as.

$$KK_{j,t+1} = (1 - dep) * KK_{j,t} + II_{j,t} \quad (2)$$

As a basic and common assumption in CGE models, each sector is a profit-maximizing firm. Therefore, in every period it employs optimal levels of intermediate inputs ($X_{j,i,t}$) and the productive factors ($F_{h,t}$), which are combined into a composite factor ($Y_{j,t}$). Although the production process occurs in one shot, in order to achieve a satisfactory understanding, this conception requires the production to be divided into two stages:

First stage:

$$\max_{F_{h,j,t}} \pi_{j,t}^y = p_{j,t}^y Y_{j,t} - \sum_h p_{h,t}^f F_{h,j,t} \quad \text{s. t.} \quad Y_{j,t} = b_{j,t} \prod_h F_{h,j,t}^{\beta_{h,j}} \quad (3)$$

Second stage:

$$\max_{Y_{j,t}, X_{j,i,t}} \pi_{j,t}^z = p_{j,t}^z Z_{j,t}^S - \left(p_{j,t}^y Y_{j,t} + \sum_j p_{j,t}^{qF} X_{j,i,t} \right) \quad (4)$$

² Iracema is a very famous native character, which also names the novel written by the nationwide known writer José de Alencar.

$$\text{s. t. } Z_{j,t}^S = \min \left(\frac{X_{j,i,t}}{ax_{j,i,t}}, \frac{Y_{j,t}}{ay_{j,t}} \right)$$

To make the description easier to understand we omit the index t . The parameter b_j is a scaling coefficient, $\beta_{h,j}$ is a share coefficient of the productive factors, the parameters $ax_{j,i}$ and ay_j are input-requirements coefficients to produce one unit of output (Z_j^S). Finally, the prices of domestic goods, composite factor, and composite good, are respectively represented by p_j^z , p_j^y , and p_j^{qF} .

2.2 Foreign Trade: the rest of Brazil and the rest of the world.

As already known, the Iracema Model represents the Ceará's economy, thus the trade must be computed as at least two different flows, which are to the rest of Brazil and to the rest of the world. Given that the economy of the Ceará state only represents around 2% of the Brazilian economy (IPECE, 2016), it is considered a small region. Combining these characteristics, namely that, an open and small economy, suggests that the Ceará state plays no significant economic impact on its trade partners. Therefore, both prices, export, and import are considered exogenous. The prices are all converted to the national currency by using a marketability margin, which in the case of foreign trade also takes into account the exchange rate.

At this part of modeling the Armington's assumption takes place, according to which all the final demanders, namely that, consumers, sectors, government, and investment goods, acquire not goods directly from the productive sector but rather the Armington's composite good (Q_i^F). It is actually a good composed by imports, coming from the rest of Brazil (M_i^C) as well as from the rest of the world (M_i^W), and local sectorial goods (Q_i^S). This combination is performed in a virtual sector for each one of the i -th Armington's composite good. The optimization problem is represented as:

$$\max_{M_i^C, M_i^W, Q_i^S} \pi_{i,t}^{qF} = p_{i,t}^{qF} Q_{i,t}^F - [p_{i,t}^{qS} Q_{i,t}^S + p_{i,t}^{mC} M_{i,t}^C + (1 + \tau^m) p_{i,t}^{mW} M_{i,t}^W] \quad (5)$$

$$\text{s. t. } Q_{i,t}^F = \gamma_i (\delta q_i^S (Q_{i,t}^S)^{\eta_i} + \delta m_i^C (M_{i,t}^C)^{\eta_i} + \delta m_i^W (M_{i,t}^W)^{\eta_i})^{\frac{1}{\eta_i}}$$

Again, the time index are omitted. Thus, the p_i^{qS} , p_i^{mC} , and p_i^{mW} are respectively the prices of the goods locally offered, imports from the rest of Brazil and the rest of the world. The following parameters are considered constant, so τ^m is an import tariff, γ_i is a scaling coefficient, and η_i is based on the elasticity of substitution. Finally, δq_i^S , δm_i^C , and δm_i^W are input share coefficients for the quantities abovementioned.

Similarly, it is necessary to analyze the decisions considering the supply side. In each period t , These decisions regards to what is offered to the intern market ($Q_{i,t}^S$) and what is exported to the rest of Brazil ($X_{i,t}^C$) as well as to the rest of the world ($X_{i,t}^W$). It is assumed that the sectors divide the domestic output by using a CET function in which every sector adjusts its output for both consumption, domestic and foreign.

$$\max_{Z_i^S, X_i^C, X_i^W, Q_i^S} \pi_{i,t}^{zS} = (p_{i,t}^{qS} Q_{i,t}^S + p_{i,t}^{xC} X_{i,t}^C + p_{i,t}^{xW} X_{i,t}^W) - (1 + \tau_i) p_{i,t}^z Z_{i,t}^S \quad (6)$$

$$\text{s. t. } Z_{i,t}^S = \theta_i \left(\xi q_i^S (Q_{i,t}^S)^{\phi_i} + \xi x_i^C (X_{i,t}^C)^{\phi_i} + \xi x_i^W (X_{i,t}^W)^{\phi_i} \right)^{\frac{1}{\phi_i}}$$

Where p_i^{xC} and p_i^{xW} represent the export prices respectively for the rest of Brazil and the rest of the world. τ_i is a joint tax composed by two taxes, namely that, the τ_i^{ICMS} for the ICMS tax and the τ_i^{OT} for all of the others taxes, including the Federal ones. It is applied on the locally produced good. Additionally, the parameter θ_i is a scaling coefficient, and ϕ_i is defined

by the elasticity of transformation. Similarly displayed in the previous maximization problem, ξq_i^S , ξx_i^C , and ξx_i^W are share coefficients.

2.3 Government

As commonly assumed in CGE models, the government is a tax collector and its consumption is set exogenous but growing at a constant rate. The government collects a direct tax from the household's income (T_H^D), which is imposed in a lump-sum style so that the fiscal balance is achieved. Additionally, it also collects from production, based on the ICMS tax (T_i^{ICMS}) and other taxes combined (T_i^{OT}) and from the imports (T_i^M). The tax rates are constant over time and expressed respectively as τ_H^D , τ_i^{ICMS} , τ_i^{OT} , and τ_i^m .

In order to cover the efficiency in the public sector, one of the main changes in the original model is the inclusion of parameters of efficiency in every channel of tax collection. Thus, the efficiency parameters associated with the tax rates are $\epsilon_{col,d}$, $\epsilon_{col,icms}$, $\epsilon_{col,ot}$, and $\epsilon_{col,m}$, respectively. Equation 11 shows the total revenue collected in a specific period t . This number already carries within it the efficiency levels coming from all of the sources of collecting.

The government spends its tax revenues in consumption of every sector (G_j^F), which is consumed in a constant ratio (μ_j), and savings (S^G). The key equations for the government are expressed by equations from (7) to (12).

$$T_{H,t}^D = \epsilon_{col,d} \tau_H^D \sum_h p_{h,t}^f F F_{h,t} \quad \forall h, t \quad (7)$$

$$T_{j,t}^M = \epsilon_{col,m} \tau_j^m p_{j,t}^m M_{j,t}^W \quad \forall j, t \quad (8)$$

$$T_{j,t}^{ICMS} = \epsilon_{col,icms} \tau_{j,t}^{ICMS} p_{j,t}^z Z_{j,t}^S \quad \forall j, t \quad (9)$$

$$T_{j,t}^{OT} = \epsilon_{col,ot} \tau_{j,t}^{OT} p_{j,t}^z Z_{j,t}^S \quad \forall j, t \quad (10)$$

$$T_{Total,t} = T_{H,t}^D + \sum_j (T_{j,t}^M + T_{j,t}^{ICMS} + T_{j,t}^{OT}) \quad \forall t \quad (11)$$

$$G_{j,t}^F \leq \frac{\mu_j}{p_{j,t}^{qF}} \left(T_{H,t}^D + \sum_j (T_{j,t}^M + T_{j,t}^{ICMS} + T_{j,t}^{OT}) - S^G \right) \quad \forall j, t \quad (12)$$

2.4 Savings and Investment

The recursive dynamics are savings-driven so that the allocation of investments among the productive sectors depends on that. The private savings in a period t (S_t^P), coming from the households, follows a propensity to save (ss^P), assumed to be constant, as indicated in the following equation:

$$S_t^P = ss^P \left(\sum_{h,j} p_{h,j,t}^f F_{h,j,t} - T_t^d \right) \quad (13)$$

The total savings is composed of the private savings plus the foreign savings, which in this paper is a combination of the difference in trading to the rest of Brazil (S_t^C) and to the rest of the world (S_t^f). This amount of resources is spent on investment goods in each sector ($I_{j,t}$) in order to accumulate capital stock. The allocation of the investment goods follows the sectoral share of operation surplus:

$$p_t^k II_{j,t} = \frac{p_{CAP,j,t}^f \zeta F_{CAP,j,t}}{\sum_i p_{CAP,i,t}^f \zeta F_{CAP,i,t}} (S_t^P + S_t^C + \varepsilon_t S_t^f) \quad (14)$$

Where the parameter ζ represents the sensitivity of investment goods allocation to the capital service price ($p_{CAP,j,t}^f$), in other words, it is a weight parameter. The parameter ε_t represents the Exchange rate.

2.5 Households

The households' aggregate consumption function is a Cobb-Douglas type and the constraint of available income composes the optimization process, specified below:

$$CC_t = a \prod_i C_{i,t}^{\alpha_i} \quad s.t. \quad \sum_i p_{i,t}^{qF} C_{i,t}^p \leq \sum_h p_{h,t}^f FF_{h,t} - S_t^p - T_{H,t}^D \quad (15)$$

Where C_i^p is the private consumption of the sector i , p_h^f and p_i^{qF} are the prices of factors and sectorial goods respectively, FF_h is the endowment of factors. Similar to the description of the sectors, the parameter a is a scaling coefficient and α_i is a share coefficient. Based on the solution coming from this problem, the fictitious objective function, or in other words, the Utility function is given as:

$$UU_t = \sum_t \frac{CC_t}{(1 + ror)^{t-1}} \quad (16)$$

2.6 Market Clearing Conditions

A CGE model describes the behavior of all the economic agents. In order to reach the general equilibrium, some assumptions are usually set. Firstly, there is no waste in this economy so that market conditions are imposed to achieve equality between supply and demand in every market. It not only means the market of goods produced by every sector, but it also includes the labor market so that the labor employed in all of the sectors must be equal to the total labor endowment, which grows at a constant rate of the population growth. Given the dynamic characteristic of the Iracema Model, the composite investment goods must be completely set in the sectors as well. Finally, Walras' Law allows us to choose a *numeraire*, one price in which all of the other prices are comparable.

3 CALIBRATION

Most of the parameters and exogenous variables required to run a CGE model are calibrated from a Social Accounting Matrix – SAM, which in turn is composed of an Input-Output table. The SAM disaggregates the Ceará's economy into six productive sectors. The disaggregation level depends on the target analysis, which in this case it is not compromised by a low level of disaggregation. This paper takes the year 2013 as the base year. The sectors considered are shown in the following table.

Table 1: Codes for short denomination of the sectors.

Code	Sectors
S1	Agriculture, including support for agriculture and post-harvest.
S2	Extractive industry.
S3	Transformation Industry, Building, Power Electricity, Water and Sewage, and Others.
S4	Trade and repair of motor vehicles and motorcycles, transportation, storage and mail, accommodation and food.
S5	Private Services.
S6	Management, security, public education and health, and social security.

Source: the authors.

This section also follows close to Hosoe et al. (2016). The calibration process in this dynamic model additionally requires information about the capital stocks employed for each sector. Once the SAM provides data for the capital services, the sectoral capital stocks are estimated by $F_{CAP,j,t} = ror * KK_{j,t}$. Combining it with the sectoral investment data, also available in the SAM, the growth rate can be observed. The deal is that the calibrated growth rate is likely to differ from the stylized fact of a growth rate generally assumed in this sort of CGE modeling known as the Business-as-Usual - BAU path, a set of assumptions that yields a steady BAU growth path at the rate of the population growth.

An adjustment in the investment registered into the SAM, given by the sum of the investments in all the sectors and formalized as $I^{SAM} = \sum_i SAM_{i,INV}$, is required to overcome the difference mentioned in the previous paragraph. The assumed investment to achieve the desired growth rate is given as $I^{Assumed} = \frac{pop+dep}{ror} * FF_{cap}^{00}$, where the parameters pop , dep , and ror are respectively the population growth rate, the depreciation rate, and the rate of returns. The FF_{cap}^{00} is the capital service input observed in the base year. Then, the ratio $adj = I^{Assumed} / I^{SAM}$ adjusts the investment good demand. It, in turn, readjust all the SAM making it compatible to a steady base equilibrium.

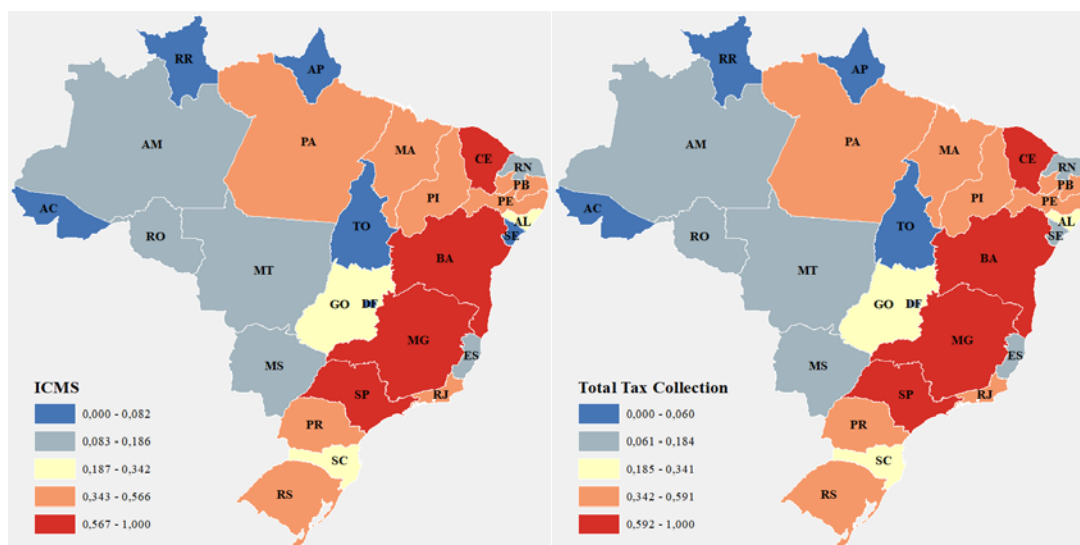
4 TAX COLLECTION EFFICIENCY IN CEARÁ STATE

This paper follows Cazals et al. (2002) and applies a nonparametric estimator known as Order- m efficiency estimation, also well explored in Tauchmann (2012), to calculate the efficiency in tax collection of both ICMS tax and the total tax collection of the Brazilian states. Although part of the analysis regards all states, the focus is the Ceará state, especially because it also works with a CGE analysis. Thus, it uses the level of efficiency performed by the Ceará state as the efficiency parameter. Even though this method is closely related to the widely known nonparametric envelopment estimators, the Data Envelopment Analysis – DEA, it is more robust in terms of extreme values and outliers.

A single input – single output approach is applied for each one of the estimations, namely that, ICMS and Total tax collection, by taking the GDP per capita as the input for both of them and the tax revenue in ICMS and the total tax revenue, which includes all the taxes the state-level government is responsible for collecting, as the outputs. We emphasize that this estimation procedure and the applied approach allow us to perform a specific analysis of efficiency besides using the estimated value for the efficiency parameter in the Iracema model. Once the biggest share of the required parameter comes from the SAM, which uses the 2013 base year, we use the same year for the efficiency parameter in order to make the calibration compatible. However, this analysis covers the period from 2012 to 2016.

As an overview of the estimations, figure 1 shows the efficiency average scores, from 2012 to 2016, for all of the Brazilian states by splitting them into five ranges of efficiency.

Figure 1: Efficiency Average (2012-2016) in ICMS collection and Total Tax Collection of the Brazilian States.

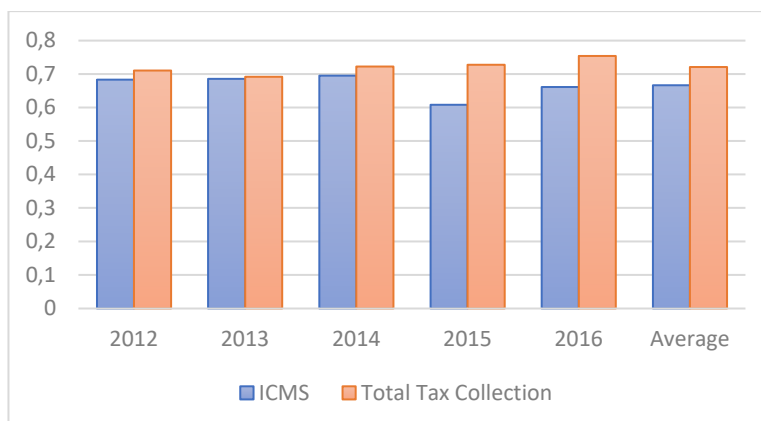


Source: The authors.

Even though this paper focuses on the Ceará state, Figure 1 is important since it provides a quick and visual comparison among the Brazilian states. Worthy to mention that although only two states show differences on the maps, improving from the lower level to the lower-intermediate level in terms of total tax collection *vis-a-vis* ICMS, namely that, the Federal District and Sergipe, they fit in a range and the real scores are actually different. The Ceará state reaches the highest range of efficiency on both measures considered, as so the states of Bahia, São Paulo, and Minas Gerais. On the other hand, the states with the poorest efficiency in tax collection are Acre, Amapá, Roraima, and Tocantins.

Now, to be more specific about the state of Ceará, Figure 2 shows the efficiency scores, year by year as well as the average, for estimations of both ICMS and Total Tax Collection. Based on that, we see it keeps constant its efficiency levels of tax collection, except for a reduction in efficiency collecting ICMS in the year 2015, although it keeps slightly above 0.60, which still fits the highest level abovementioned. Meanwhile, the efficiency of total tax collection presents a trend of increasing.

Figure 2: Efficiency Level in Collection of ICMS and Total Taxes for the Ceará state.



Source: The authors.

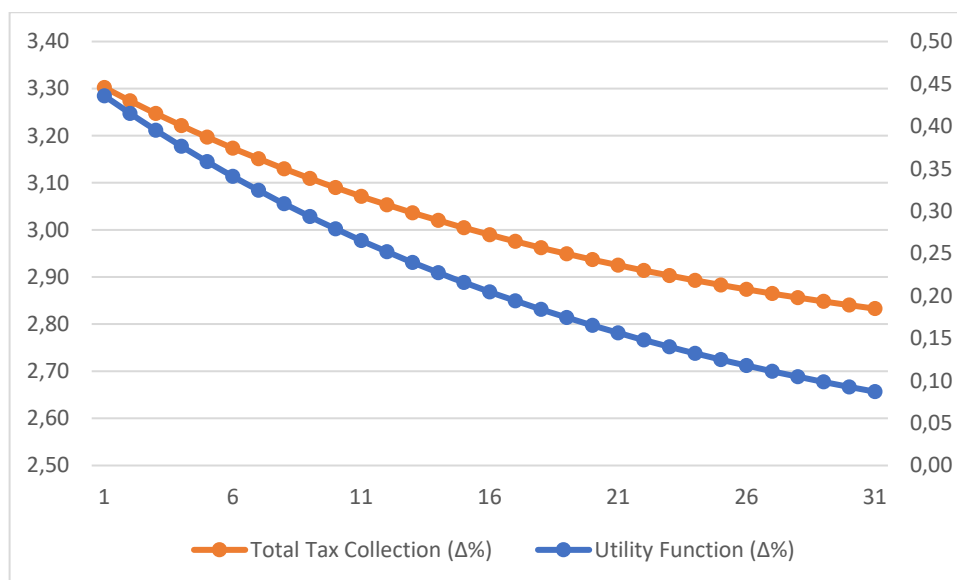
5 SIMULATION EXERCISES AND RESULTS

This paper concerns public sector efficiency issues in tax collection. Therefore, using the Iracema Model it performs simulation exercises by changing the efficiency levels on taxation for the Ceará state. The assumption behind it is that the government enhances its efficiency by collecting taxes. Although efficiency parameters are included in all the sources of tax collection and also that the analysis in the previous section contemplates efficiencies on both ICMS and Total tax collection, aiming to achieve more accurate outputs, the exercises are more specific and simulate changes in efficiency only in ICMS tax. An additional reason for this specific approach is the importance of this tax for the Brazilian states and the control they have over it. Once the Calibration process is based on data for the year 2013, the efficiency level considered refers to the same year, which is 0.685.

5.1 Improving Efficiency in Tax Collection of ICMS.

Firstly, the simulation in this section consists of a positive variation of 10% over the efficiency level of the ICMS collection, which is changing from the base year value of 0.685 to 0.753. This change is moderate and feasible. The concerns on the tax revenue of the Ceará state already expressed in the introduction section requires that the total tax collection changes be contemplated in the first analysis. Besides, considering the common measure usually delivered as a summary of the results in a CGE model, the percentage variation of the utility function, based on consumption, represents the variations in wellbeing. Thus, figure 2 shows the evolution of these two measures over the 30 periods simulated.

Figure 3: **Percentage Change of the Total Tax Collection and Utility Function from a 10% Improvement of Efficiency Collecting ICMS tax.**



Source: the authors.

On the left axis, there is a total tax collection, which represents a 3.3% variation in the first period and a 2.83% in the last. Given it regards the total tax revenue of the Ceará state, these values can represent a significant volume of resources, especially considering that the efficiency shocks are performed in only one sort of tax rate. On the other side, the right axis shows the percentage change in utility. It starts achieving a 0.44% change and finishes with an insignificant 0.09%, but still positive. The decreasing trend over the years regards the way the variables evolve, considering the population growth rate.

The model applied has a plethora of variables hence the analysis must pick just some of them. Table 2 below contemplates eight variables, which are the Tax Collection of ICMS (T^{ICMS}), the total tax collection of the sector j (T_j), the private consumption (C^p), the GDP (Z), the exports for the rest of the world (Xw) and the rest of Brazil (Xc), and the imports for the rest of the world (Mw) and the rest of Brazil (Mc). All the variations are represented in terms of percentage variations (%). Given the dynamics of the model, it is worth to mention the denomination of the periods. Four periods are displayed, namely that, the first, the fifth, the tenth, and the twentieth, for all sectors. It is worth mentioning that, for simplicity, it is assumed that the efficiency improvements reach all the sectors equally.

Table 2: Percentage Change of Selected Variables from a 10% Increase in Efficiency Tax Collection of ICMS.

Sector/Period	ΔT^{ICMS}	ΔT_j	ΔC^p	ΔZ	ΔXw	ΔMw	ΔXc	ΔMc
S1P1	9.91	2.64	0.470	-0.053	-0.196	0.128	-0.150	0.082
S1P5	9.78	2.53	0.367	-0.206	-0.437	0.091	-0.381	0.035
S1P10	9.66	2.42	0.260	-0.366	-0.689	0.055	-0.624	-0.011
S1P20	9.47	2.26	0.103	-0.603	-1.066	0.006	-0.990	-0.071
S2P1	9.73	6.35	0.535	-0.145	-0.149	-0.254	-0.103	-0.300
S2P5	9.53	6.16	0.462	-0.339	-0.373	-0.360	-0.317	-0.416
S2P10	9.31	5.95	0.386	-0.557	-0.626	-0.464	-0.561	-0.529
S2P20	8.94	5.59	0.270	-0.920	-1.055	-0.604	-0.978	-0.681
S3P1	9.68	2.26	0.502	-0.221	-0.285	-0.132	-0.239	-0.178
S3P5	9.56	2.16	0.423	-0.341	-0.439	-0.193	-0.383	-0.249
S3P10	9.44	2.05	0.342	-0.464	-0.595	-0.257	-0.530	-0.323
S3P20	9.27	1.90	0.226	-0.638	-0.814	-0.351	-0.737	-0.428
S4P1	10.07	4.21	0.357	0.014	-0.294	0.409	-0.248	0.363
S4P5	9.99	4.13	0.281	-0.070	-0.396	0.347	-0.340	0.291
S4P10	9.90	4.05	0.204	-0.155	-0.498	0.283	-0.433	0.217
S4P20	9.78	3.93	0.094	-0.275	-0.639	0.186	-0.563	0.108
S5P1	10.22	8.01	0.324	0.086	-0.355	0.532	-0.309	0.486
S5P5	10.15	7.93	0.262	0.023	-0.409	0.460	-0.353	0.403
S5P10	10.07	7.86	0.198	-0.041	-0.463	0.385	-0.398	0.319
S5P20	9.95	7.75	0.107	-0.134	-0.539	0.276	-0.462	0.198
S6P1	10.00	9.52	0.424	-0.002	-0.216	0.211	-0.170	0.165
S6P5	9.95	9.47	0.399	-0.004	-0.137	0.129	-0.081	0.073
S6P10	9.90	9.42	0.373	-0.006	-0.054	0.043	0.012	-0.023
S6P20	9.83	9.35	0.338	-0.008	0.071	-0.087	0.149	-0.164

Source: The authors. Note: All of the variations represented in the table are percentage variations (%).

As expected, the results show increases in tax collection in ICMS and in the total collected from each sector j at a lower level since ICMS is just a share of it. Additionally, the decreasing trend, already mentioned, is a pattern observed for all variables, given that the population growth is being computed over the years.

However, the response to the shock is different for each sector. Sectors 4 and 5, for instance, start the implementation period with an ICMS collection higher than 10%, which represents the shock. Also, the total collected from a sector depends on the share the ICMS represents on it. For instance, sectors 3 and 1 present the lowest impacts on this variable, with respectively 2.26% and 2.64%. Conversely, sectors 6 and 5 present the highest impacts on the total collected from a sector, with respectively a 9.5% and an 8% increase. The results for the

consumption, although seem no significant, are responsible for the changes in wellbeing previously mentioned.

5.2 Adapting the Double Dividend Hypothesis: Efficiency and Tax Rates.

As already mentioned in the introduction, the tax burden in Brazil is high and so it is when it concerns state-level taxes. Although this paper has already shown that both the central and the state-level governments are facing an upcoming period of short and limited resources due to the current scenario, there might still be some margin to apply tax reductions, probably not in a neat way but by applying some sort of compensation.

This compensation could be generated by adapting the double dividend hypothesis and using the recycling revenue issues. The double-dividend hypothesis, from Fullerton and Metcalf (1997), states that create or raise taxes on polluting activities might be able to provide two types of benefits. The first regards an improvement in the environment itself by reducing pollution and the second concerns an improvement in the efficiency of the economy by using the tax revenues collected by the new or higher taxes to reduce other taxes, which is also known as recycling revenues. However, it emphasizes that the validity of this hypothesis must not be taken as applicable in a general matter. Following this specific issue, after performing simulations, Freire-Gonzalez (2018), claims that even though the environmental benefit is almost always achieved due to the fact the changes in taxes also changes the behavior of the economic agents, there is no consensus in the literature, suggesting that it still needs further research.

In this literature, both Goulder and Hafstead (2013) and Beck et al. (2015) apply general equilibrium models, the former for the United States and the latter for Canada. Goulder and Hafstead (2013) evaluate what alternative tax reductions could be financed by the revenues from an environmental tax. They found that this tax promotes a substantial net revenue and also that the impacts on GDP and wellbeing strongly depend on how this revenue is recycled to benefit the productive sectors.

Roughly speaking, assuming that creating or increasing some tax generates higher tax collection allowing to reduce some other, this process can seem as a composition of revenues. Therefore, it might be possible to compensate the components of a tax collection taking into account the efficiency by collecting. The following axiom formalizes this idea.

Axiom: Compensation of the Tax Collection Drivers.

Let the equation for the Tax Collection be $T_{j,t}^i = \epsilon_{col,t}^i \tau_{j,t}^i Z_{j,t}^i$. From an i type of tax, a sector j , and at time t , the components, namely that, $\epsilon_{col,t}^i$, $\tau_{j,t}^i$, and $Z_{j,t}^i$ representing respectively the efficiency level on tax collection, the tax rate, and the incidence base of them, are the drivers of the tax collection $T_{j,t}^i$. For a given level of efficiency, higher than the current value, there is at least one value for the tax rate, lower than the current value, able to yield either the same level or a higher level of the Tax Collection.

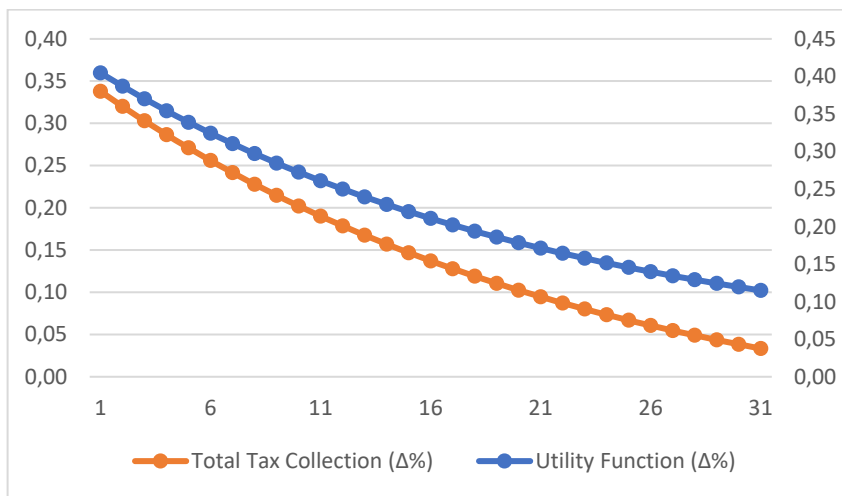
The axiom is generally set. Therefore, once included in a CGE model, it is supposed to be able to produce similar conclusions in other variables such as the total tax collection of a government and the wellbeing. The following simulation exercise, using the IRACEMA model proposed in this paper, works as an empirical demonstration of it.

Keeping the modest change in the efficiency of 10% as performed in the previous section and applying the axiom of the compensation of the tax collection drivers to keep the total tax revenues unchanged but still positive over the 30 periods considered, we found that it

needs an 8.75% reduction on tax rates. It might be related to the boost effects this sort of incentives bring to the productive sectors.

Due to the population growth inserted on the variation, there is a trend of diminishing effects over the years. So, the combination is chosen in a way there are no negative effects over all the years. This is the reason there are still some positive effects in the beginning. Worthy to mention that the changes in Total collection and Well-being are both lower than 0.4% in the first period after the implementation, as observed in figure 4. Based on that one can assume the percentages chosen fits the axiom.

Figure 4: Percentage Change of the Total Tax Collection and Utility Function from a Combination of a 10% Improvement of Efficiency Collecting ICMS tax and an 8.75% reduction in ICMS tax rate.



Source: the authors.

Similar to the previous analysis, the percentage variation of total tax collection is placed in the left axis and it is around 0.34% in the first period and of 0.03% in the last. On the right axis, there is a percentage change in utility starting at 0.4% and finishing at 0.12%.

Table 3: Percentage Change of Selected Variables from a Combination of a 10% Improvement of Efficiency Collecting ICMS tax and an 8.75% reduction in ICMS tax rate.

Sector/Period	ΔT_{icms}	ΔT_s	ΔC^p	ΔZ	ΔX_w	ΔM_w	ΔX_c	ΔM_c
S1P1	0.325	0.101	0.436	-0.077	-0.361	0.316	-0.382	0.337
S1P5	0.204	-0.011	0.313	-0.277	-0.727	0.342	-0.838	0.370
S1P10	0.058	-0.146	0.167	-0.517	-1.162	0.371	-1.199	0.409
S1P20	-0.110	-0.301	-0.005	-0.796	-1.665	0.402	-1.712	0.450
S2P1	0.356	0.248	0.540	-0.127	-0.165	0.074	-0.185	0.095
S2P5	0.177	0.069	0.475	-0.364	-0.390	0.034	-0.419	0.062
S2P10	-0.060	-0.168	0.395	-0.575	-0.695	-0.015	-0.732	0.022
S2P20	-0.385	-0.492	0.296	-0.932	-1.123	-0.068	-1.171	-0.020
S3P1	0.542	0.274	0.590	0.084	0.183	-0.100	0.162	-0.080
S3P5	0.476	0.194	0.527	0.024	0.128	-0.162	0.099	-0.133
S3P10	0.396	0.128	0.475	-0.051	0.056	-0.235	0.019	-0.198
S3P20	0.300	0.034	0.398	-0.143	-0.034	-0.321	-0.082	-0.274
S4P1	0.460	0.275	0.195	-0.106	-0.725	0.722	-0.746	0.743
S4P5	0.392	0.192	0.075	-0.208	-0.902	0.722	-0.930	0.751
S4P10	0.311	0.127	-0.015	-0.324	-1.095	0.713	-1.132	0.750
S4P20	0.219	0.035	-0.135	-0.450	-1.291	0.684	-1.338	0.732
S5P1	0.816	0.745	0.282	0.049	-0.532	0.638	-0.552	0.659
S5P5	0.755	0.684	0.214	-0.010	-0.593	0.583	-0.622	0.611
S5P10	0.682	0.611	0.160	-0.079	-0.664	0.515	-0.701	0.552
S5P20	0.599	0.528	0.083	-0.158	-0.740	0.432	-0.787	0.480
S6P1	0.730	0.716	-0.029	-0.009	-1.162	1.158	-1.183	1.179
S6P5	0.691	0.677	-0.051	-0.011	-1.099	1.090	-1.127	1.119
S6P10	0.645	0.631	-0.075	-0.012	-1.020	1.006	-1.057	1.044
S6P20	0.590	0.576	-0.103	-0.014	-0.920	0.901	-0.968	0.949

Source: the authors.

Reducing ICMS, which is the main tax for state tiers of governments, unlike the previous analysis this exercise might shake the economy since there are venues to spread the impacts of this shock making them reverberate both directly and indirectly on a set of different variables. Although in some periods is a negative change in consumption, the aggregated net effect is positive, as seen in figure 4. Remembering the focus here is to guarantee that the Total tax collection seems unchanged, but in a way it still marginally positive, over 30 periods. This type of compensation can surely be applied controlling for any other variable present in the model such as the consumption and/or the tax collected from each sector.

The simulation presented, in which the compensation reaches a minimum positive gain, is some sort of an extreme case. Remembering that the axiom also fits for different levels of compensations that could increase the tax collection. As already shown along the results presented so far, the efficiency gains itself increases the collection, *ceteris paribus*. Then, reducing taxes but lower than the level required to exhaust the gains coming from the efficiency gains, there will still gather a higher amount collected. Besides, this process is a booster effect in productive sectors provided by lower tax rates. These types of outcomes support the original proposition of Laffer's curve according to which tax rates should be reduced to increase the tax collection efficiency (CEVIK ET AL., 2019).

6 FINAL REMARKS AND POLICY IMPLICATIONS.

The general contribution of this paper is to analyze the efficiency in tax collection as an alternative to overcome or mitigate residual problems on balance the budget in Ceará state left by the Covid-19 pandemic. It required a double analysis, firstly estimating the efficiency in tax collection and secondly adapting a CGE model to consider these important issues. Finally, applying the values estimated into the model, exercises were performed simulating improvements in efficiency levels and also combining it with tax rates reductions.

The current fiscal situation of the Brazilian central government, which was already facing problems to balance its budget even before the Covid-19 pandemic starts and then with all resources it has transferred, plus the damage it has caused in the economic activities, constitutes a strong unfeasibility for raise taxes in the current economic and political scenario. Once tax changes require higher levels of political capital, one would say that it is not a feasible solution to recover public finances.

Conversely, it is worth mentioning that nowadays the alternative proposed in this paper, improve tax collection efficiency, is feasible. It can be performed, for instance, by improving the collecting system technologically such as using electronic invoices. Additionally, according to the Subnational Entities Finance Bulletin, STN (2020), there are possibilities to review significant reducers of collection, namely that, tax waiver and tax expense, which are indirect government expenditures made through the tax system reducing potential revenue. Therefore, there is the possibility to perform continuous reassessments of them and convert them into higher tax collection. It would be a solution to the problem pointed out in this paper.

Moreover, as long as these measures can potentially reverse the effects of the Covid-19 pandemic on the public finances or at least mitigate them, in normal economic scenarios they can also be applied to improve the business environment and boost the economy as development-oriented policies. We emphasize replicability as an important feature of these policies. Therefore, the results shown for the Ceará state might be observed in other subnational units or even in terms of the central government, considering it is responsible for collecting a significant part of tax revenues.

Given that this sort of policy simulated along this paper has the potential to enrich the tax collection and that the effects reverberate for some periods, once the negative effects from the Covid-19 pandemic have been surpassed, depending on the values, there will still be some positive effects. It might allow other uses for the extra resources such as implement a reduction on the public debt or even increase public investments, which might work as extra positive incentives for the productive sectors, similar to lowering tax rates.

The exercises performed on this paper as well as the results presented might assist the policymakers to enrich their understandings helping them to find out other alternatives in which they can base new public policies. We emphasize that the specificities of the unit and the current economic environment must be taken into account. The most important takeaways are the feasibility and the replicability of these improving efficiency policies.

Finally, although the Iracema model contributes itself to the literature by including efficiency on tax collection in a dynamic and regional CGE model, further extensions of it might be considered. One example is the possibility to split the households into different categories to capture distributive effects and/or include consumption of public goods into the Utility function. It would allow us to include efficiency in the provision of public goods.

Moreover, for further improvements specifically on the efficiency part, performing regressions to verify exogenous drivers for the tax efficiency would be useful for achieve a deeper understanding and orientate the designs of public policies as additional efforts to improve the efficiency of tax collection. This approach, from Simar and Wilson (2007), as a step further of the estimations of efficiency, it performs regressions by taking the estimated scores as the dependent variables.

REFERENCES

- AIZENMAN, O. JINJARAK, Y. The collection efficiency of the Value Added Tax: Theory and international evidence. **Journal of International Trade and Economic Development**, v. 17, n. 3, p 391-410, 2018. DOI: 10.1080/09638190802137059
- ARBEL, A. FIALKOFF, C. KERNER, A. Public policy for reducing tax evasion: implications of the Yule–Simpson paradox, **Applied Economics Letters**, v. 26, n. 13, p. 1092-1099. 2019. DOI: 10.1080/13504851.2018.1537471
- ASHRAF, B. Economic impact of government interventions during the COVID-19 pandemic: International evidence from financial markets. **Journal of Behavioral and Experimental Finance**, v. 27. 2020. <https://doi.org/10.1016/j.jbef.2020.100371>
- ATKESON, A. What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios. **NBER Working Paper No. 26867**. March 2020
- BECK, M. RIVERS, N. WIGLE, R. YONEZAWA, H. Carbon tax and revenue recycling: Impacts on households in British Columbia. **Resource and Energy Economics**, v. 41, p. 40–69. 2015. <http://dx.doi.org/10.1016/j.reseneeco.2015.04.005>
- BRAATZ, J. GONÇALVES, R. PINTO, G. MORAES, G. Proposta de reestruturação tributária e orçamentária para o RS: uma análise em Equilíbrio Geral Computável. **Perspectiva Econômica**, v. 11, n. 2, p. 95-114, 2015.
- BUHEJI, M. CUNHA, K. BEKA, B. MAVRIĆ, B. SOUZA, Y. SILVA, S. HANAFI, M. YEIN, T. The Extent of COVID-19 Pandemic Socio-Economic Impact on Global Poverty. A Global Integrative Multidisciplinary Review. **American Journal of Economics**, v. 10, n. 4, p. 213-224. 2020. DOI: 10.5923/j.economics.20201004.02
- CAZALS, C. FLORENS, J.P. SIMAR, L. Nonparametric frontier estimation: a robust approach. **Journal of Econometrics**, v. 106, n. 98, p. 1–25, 2002.
- CEVIK, S. GOTTSCHALK, J. HUTTON, E. JARAMILLO, L. KARNANE P. SOW, M. Structural Transformation and Tax Efficiency. **IMF Working Paper WP/19/30**. Washington D.C., Feb., 2019.
- DIXON, P. PARMENTER, B. SUTTON, J. Spatial disaggregation of ORANI results: A preliminary analysis of the impacts of protection at the state level. **Econ. Anal. Policy**, v. 8, p. 35-86. 1978.
- DIXON, P. PARMENTER, B. SUTTON, J. VINCENT, D. **ORANI: A Multisectoral Model of the Australian Economy, Contributions to Economic Analysis**, North Holland Publishing Company. 1982.

- DIXON, P. PARMENTER, B. Computable general equilibrium modeling for policy analysis and forecasting. In: H. M. Amman, D. A. Kendrick, et al. **Handbook of computational economics**. Amsterdam; New York: Elsevier. 1996.
- DOMINGUES, E. HADDAD, E. Analyzing the spatial impact of tax policies: An interregional CGE framework for Brazil, paper presented at the **Policy Modeling International Conference**, Brussels, Belgium, 4-6 July. 2002.
- DOMINGUES, E. P.; LEMOS, M. Regional impacts of trade liberalization strategies in Brazil, paper presented at the **Input-Output and General Equilibrium Data, Modeling and Policy Analysis Conference**, Brussels, Belgium, 2-4 September. 2004.
- DOMINGUES, E. P. CARDOSO, D. F. AND MAGALHÃES, A. S. A pandemia do Corona vírus no Brasil: demanda emergencial de setores relacionados a saúde e impactos econômicos. **Working Paper**, NEMEA – Núcleo de Estudos em Modelagem Econômica e Ambiental Aplicada do Cedeplar – UFMG, Belo Horizonte, MG. 2020.
- DOMINGUES, E. GUEDES, G. RIBEIRO, R. MAGALHÃES, A. FREIRE, D. SANTOS, R. FELIX, M. ANDRADE, J. SIMONATO, T. MIYAJIMA, D. Cenários de isolamento social da COVID19 e impactos econômicos em Minas Gerais. **Nota Técnica**. NEMEA – Núcleo de Estudos em Modelagem Econômica e Ambiental Aplicada do Cedeplar – UFMG, Belo Horizonte, MG. 2020.
- FERNANDES, N. Economic Effects of Coronavirus Outbreak (COVID-19) on the world economy. **Working Paper**. IESE Business School, Spain. Apr., 2020.
- FOCHEZATTO, Adelar. Testando um modelo de equilíbrio geral computável para a economia gaúcha: impactos da reestruturação tributária. **Ensaio FEE**, v. 23, p. 371-398, 2002.
- FREIRE-GONZALEZ, J. Environmental taxation and the double dividend hypothesis in CGE modelling literature: A critical review. **Journal of Policy Modeling**, v. 40, p. 194–223. 2018. <https://doi.org/10.1016/j.jpolmod.2017.11.002>
- FULLERTON, D. METCALF, G. E. Environmental Taxes and the Double-Dividend Hypothesis: Did You Really Expect Something for Nothing? **Working Paper 6199**, NBER WORKING PAPER SERIES. NATIONAL BUREAU OF ECONOMICS RESEARCH. Sept., 1997.
- GOULDER, L. H., HAFSTEAD, M. A. C. Tax Reform and Environmental Policy: Options for recycling revenue from a tax on carbon dioxide. **Discussion Paper - Resources for the Future**, RFF DP 13-31. Oct., 2013.
- GUILHOTO, J. J. M. **Um modelo computável de equilíbrio geral para planejamento e análises de políticas agrícolas (PAPA) na economia brasileira**. (Tese de Livre-Docência). ESALq, Universidade de São Paulo, Piracicaba, 1995.
- HADDAD, E. A. **Regional inequality and structural changes: Lessons from the Brazilian experience**. Aldershot: Ashgate, 1999.
- HADDAD, E. A. **Retornos crescentes, custos de transporte e crescimento regional**. São Paulo: USP, 2004. Tese de Livre-Docência. Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, 2004.

HADDAD, E. PEROBELLI, F. Trade liberalization and regional inequity: Do transportation costs impose a spatial poverty trap? **EcoMod2005 International Conference on Policy Modeling**, 29 June-1 July. 2005.

HORRIDGE, M. MADDEN, J. WITWER, G. Using a highly disaggregated multiregional single-country model to analyze the impacts of the 2002-3 drought on Australia, paper presented at the **2003 GTAP Conference**, Netherlands, June, 2003.

HOSOE, N., GASAWA, K., & HASHIMOTO, H. **Textbook of computable general equilibrium modelling: programming and simulations**. Palgrave Macmillan. 2010.

HOSOE, N. GASAWA, K. HASHIMOTO, H. **Textbook of Computable General Equilibrium Modeling: Programming and Simulations**, 2nd Edition, University of Tokyo Press. 2016.

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ - IPECE. **Ceará em Números 2016**. Secretaria de Planejamento e Gestão, Governo do Estado do Ceará. Fortaleza, Ceara, Brasil. 2016.

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ – IPECE. **PIB Trimestral do Ceará: 2º Trimestre**. Setembro/2020. Fortaleza, Ceará, Brasil. 2020a.

Available at: https://www.ipece.ce.gov.br/wp-content/uploads/sites/45/2020/09/APRESENTACAO_PIB_2o_TRIM2020.pdf

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ – IPECE. Queda na Arrecadação de ICMS e Auxílio Emergencial. **Enfoque Econômico N° 216**, Outubro/2020. Fortaleza, Ceará, Brasil. 2020b.

Available at: https://www.ipece.ce.gov.br/wp-content/uploads/sites/45/2020/05/EnfoqueEconomicoN216_12_05_2020.pdf

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ – IPECE. Os principais resultados da PNAD - COVID mensal do IBGE para o Ceará. **IPECE Informe N° 180**, Setembro/2020. Fortaleza, Ceará, Brasil. 2020c.

Available at: https://www.ipece.ce.gov.br/wp-content/uploads/sites/45/2020/10/ipece_informe_180_01_out2020.pdf

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ – IPECE. Breve Análise dos Impactos da Crise Sanitária nas Principais Fontes de Financiamento do Estado do Ceará e das Transferências Compensatórias. **IPECE Informe N° 182**, Outubro/2020. Fortaleza, Ceará, Brasil. 2020d.

Available at: https://www.ipece.ce.gov.br/wp-content/uploads/sites/45/2020/10/ipece_informe_182_20_out2020.pdf

INSTITUIÇÃO FISCAL INDEPENDENTE - IFI. **Relatório de Acompanhamento Fiscal N° 25**. Outubro, 2020.

Available at: <https://www2.senado.leg.br/bdsf/handle/id/578343>.

LUCIO, F. G. C. PAIVA, W.L. PEREIRA, R. A. C. PENNA, C. M. DALL'ERBA, S. Public Sector Overall Efficiency: an application in the Ceará state. In **Anais do XV Encontro Economia do Ceará em Debate**. Fortaleza – Ceará. Novembro, 2019. Available at: http://www2.ipece.ce.gov.br/encontro/2019/artigos_aprovados/PUBLIC%20SECTOR%20

VERALL%20EFFICIENCY_%20AN%20APPLICATION%20IN%20THE%20CEAR%C3%81%20STATE_%20BRAZIL.pdf

MARTIN, A. MARKHVIDA, M. HALLEGATTE, S. WALSH, B. Socio-Economic Impacts of COVID-19 on Household Consumption and Poverty. **Economics of Disasters and Climate Change**, v. 4, p. 453–479, 2020.

MCKIBBIN, W. J. FERNANDO, R. The Global Macroeconomic Impacts of COVID-19: Seven Scenarios. **CAMA Working Paper No. 19/2020**, Centre for Applied Macroeconomics Analysis. The Australian National University. Mar., 2020. Available at SSRN: <https://ssrn.com/abstract=3547729>

MORRISSEY, O. VON HALDENWANG, C. VON SCHILLER, A. IVANYNA, M. BORDON, I. Tax Revenue Performance and Vulnerability in Developing Countries, **The Journal of Development Studies**, v. 52, n. 12, p. 1689-1703. 2016.
DOI: 10.1080/00220388.2016.1153071

PORSSE, A. A. **Competição tributária regional, externalidades fiscais e federalismo no Brasil**: uma abordagem de equilíbrio geral computável. Porto Alegre, UFRGS, 2005. Tese (Doutorado em Economia) Faculdade de Ciências Econômicas, Universidade Feredal do Rio Grande do Sul, 2005.

RIBEIRO, L. C. S. **Investimentos estruturantes e desigualdades regionais na região nordeste do brasil**. Editora UFS. São Critóvão, 2017.

ROMER, C. D. ROMER, D. H. The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. **American Economic Review**, v. 100, 763–801, 2010.

SECRETARIA DO TESOURO NACIONAL – STN. **Boletim de Finanças dos Entes Subnacionais**. Outubro, 2020.
Available at: <https://www.tesourotransparente.gov.br/publicacoes/boletim-de-financas-dos-entes-subnacionais/2020/114>.

SIMAR, L. WILSON, P. W. Estimation and inference in two-stage, semi-parametric models of production processes. **Journal of Econometrics**, v. 136, n. 1, p. 31–64, 2007.
<https://doi.org/10.1016/j.jeconom.2005.07.009>

SUMNER, A. HOY, C. ORTIZ-JUAREZ, E. Estimates of the impact of COVID-19 on global poverty. **WIDER Working Paper 2020/43**. United Nations University World Institute for Development Economics Research. Apr., 2020.

TAUCHMANN, Harald. Partial frontier efficiency analysis. **The Stata Journal**, v. 12, n. 3, p. 461–478, 2012.

ZHANG, D. HU, M. AND JI, Q. Financial Markets under the global pandemic of COVID-19. **Finance Research Letters**, 2020. <https://doi.org/10.1016/j.frl.2020.101528>.

MODELO DE EQUILÍBRIO GERAL COMPUTÁVEL E ANÁLISE INSUMO-PRODUTO: CONTRASTANDO APLICAÇÕES PARA O CEARÁ

Christiano Modesto Penna*

Ricardo de Castro Pereira**

Nicolino Trompieri Neto***

Witalo de Lima Paiva****

RESUMO

Há uma grande variedade de modelos para se avaliar políticas econômicas e seus impactos em âmbito regional. Especificamente para a Economia do Ceará, o IPECE vem trabalhando com Análises Insumo-Produto (AIO) e com um Modelo de Equilíbrio Geral Computável (CGE). Embora essas duas classes de modelos venham sendo amplamente utilizadas, as bases teóricas destes modelos costumam ser bastante distintas, o que pode gerar resultados até mesmo conflitantes ao se conduzir exercícios empíricos sob essas perspectivas. No Brasil, nenhum estudo comparou as previsões de diferentes modelos para uma mesma área geográfica usando-se um mesmo conjunto de dados. Dado que tanto a AIO padrão quanto o modelo CGE do IPECE (MARES/CE) utilizam como fonte de dados a Matriz Insumo-Produto (MIP) e a Tabela de Recursos e Usos (TRU) do ano de 2013, a comparação entre as previsões geradas por essas duas metodologias é bastante valiosa, tanto do ponto de vista científico, quanto do ponto de vista político, já que a magnitude e a maneira como os impactos são disseminados para o restante da economia cearense tendem a diferir entre as modelagens em questão. Tendo isso em vista, o artigo busca contrastar as estruturas teóricas e as previsões empíricas dessas duas metodologias.

1 INTRODUÇÃO

De modo geral, uma Avaliação de Impacto Econômico (AIE) é uma análise quantitativa. De modo específico, ela é uma projeção a priori de como um choque afetará os indicadores econômicos de uma determinada região.

É recorrente se fazer Avaliações de Impacto Econômico em âmbito regional com base em Análises de Insumo-Produto (AIO) e Modelos de Equilíbrio Geral Computável (CGEs).¹ O Estado do Ceará segue a comunidade acadêmica e conduz análises também embasadas nessas ferramentas. Os impactos econômicos do Aquário do Ceará, por exemplo, foram analisados pela ótica Insumo-Produto [ver Paiva e Trompieri Neto (2015)]. Em contraponto, Paiva et al

* Prof. do CAEN/UFC – E-mail: cmp@caen.ufc.br. Tel.: (85)997366006.

** Doutor em Economia pela Escola Brasileira de Economia e Finanças (EPGE) da Fundação Getúlio Vargas (FGV – RJ). Diretor de Estudos econômicos (DIEC) - IPECE e professor associado do Departamento de Teoria Econômica da UFC e do CAEN/UFC.CAEN/UFC.

*** Doutorado em Economia CAEN/UFC e Analista De Políticas Públicas – IPECE.

**** Doutorado em Economia CAEN/UFC e Analista De Políticas Públicas – IPECE.

¹ Há, ainda, modelos ditos Híbridos, os quais reúnem elementos de econometria e de Matriz Insumo-Produto [ver, por exemplo, Montoya e Pasqual (2015) e Firme e Perobelli (2012)].

(2018) apresentam o MARES/CE, que é um modelo CGE desenvolvido pelo IPECE, e conduzem dois exercícios empíricos para mostrar a aplicabilidade do modelo. Mas especificamente, são conduzidas análises onde se simulam os efeitos i) de redução de 5% nas transferências do governo para as famílias e, ii) de redução na alíquota do ICMS sobre o setor da Agricultura.

Embora os modelos AIO e CGE sejam amplamente utilizados na avaliação de impactos, uma comparação entre seus resultados só foi feita em nível teórico [ver, por exemplo, [Rose \(1995\)](#); [Rose \(2004\)](#); [Okuyama e Santos 2014](#)] ou com base em diferentes estudos de caso [ver, por exemplo, [Okuyama \(2010\)](#)]. Existem poucos estudos em que ambos os tipos de modelos são comparados empiricamente de forma sistemática para o mesmo estudo de caso e área geográfica, e usando uma base de dados idênticas [[Hu et al. \(2014\)](#); [West \(1995\)](#)].

Tal comparação é oportuna, tanto do ponto de vista científico quanto político, pois a magnitude e a distribuição espacial-econômica do impacto podem diferir drasticamente entre os modelos. É possível, por exemplo, que os investimentos em um determinado setor sejam justificados por um determinado modelo, mas completamente desaprovados por outro. Já a maneira como os impactos diretos e indiretos de determinada política pode se disseminar para o restante da economia também podem ser diferentes, dependendo da escolha do modelo. Ademais, as estimativas agregadas desse impacto também devem ser distintas, dadas as bases teóricas sob as quais os modelos são construídos.

Nesse estudo contrastamos as bases teóricas da AIO padrão com a teoria por trás do modelo CGE do IPECE (MARES/CE). O objetivo é averiguar se essas modelagens corroboram e/ou contradizem uma à outra em termos de teoria e evidências empíricas.

Por questões de exiguidade textual, descrever toda a metodologia por trás das abordagens AIO e CGE tornou-se inviável, no entanto, fica como sugestão de apoio o livro texto de [Miller e Blair \(2019\)](#), para adentrar na AIO, e o artigo de [Paiva et al \(2018\)](#), onde se discute o MARES/CE.

O trabalho prossegue da seguinte forma: na próxima seção se discute as bases teóricas das abordagens AIO e CGE. A Seção 3 toma proveito da teoria e busca averiguar as predições empíricas desses modelos. Mais especificamente, comparam-se indicadores de encadeamento (AIO) e produtividade (MARES/CE), realiza-se uma simulação com tecnologia estática (ampliando-se exógenamente o investimento), assim como se contrasta a abordagem do campo de influência (AIO) com um exercício contrafactual (MARES/CE). Também se apresentam os dados usados com esse intuito. Por fim, a Seção 4 conclui o trabalho ressaltando suas limitações e possíveis avanços, além de fornecer lições e recomendações para profissionais e formuladores de políticas regionais/setoriais.

2 BASES TEÓRICAS: AIO VS CGE

Em geral, a AIO padrão pode ser descrita por um modelo estático linear que trata a economia por meio de um conjunto de inter-relações setoriais (firmas) e finais (consumidores). Um modelo CGE neoclássico, por outro lado, é um sistema de equações em que a concorrência perfeita é assumida no mercado de produtos e as dotações de fatores são totalmente empregadas. Em cada região, uma empresa representativa maximiza os lucros para uma dada restrição tecnológica, e uma família representativa maximiza a utilidade do consumo sob uma restrição orçamentária. O fechamento macroeconômico é neoclássico, ou seja, os investimentos são determinados pela poupança, sendo que a demanda por fatores de produção é igual à sua oferta (fixa). Uma proporção fixa da renda familiar é destinada à poupança e há um agente hipotético que coleta todas as poupanças e as direciona para investimentos. Além disso, a balança comercial é determinada endogenamente, e não dada exógenamente como nos modelos IO.

Os modelos IO e CGE são caracterizados por uma série de diferenças. A Tabela 1, a seguir, foi adaptada a partir de [Koks et al \(2016\)](#) e reúne as diferenças mais marcantes.

TABELA 1: DIFERENÇAS ENTRE MODELAGENS AIO E CGE:

Característica	AIO	CGE
Horizonte de tempo	Curto prazo	Longo prazo
Substituição de insumos	Não há possibilidade	É possível
Complexidade matemática	Linear / Simples	Não linear / Avançada
Interdependência setorial	Decorre de coeficientes técnicos	Decorre de elasticidades cruzadas
Efeito transbordamento	Ocorre pelo lado da demanda apenas	Ocorre pelo lado da oferta e da demanda
Lado da oferta	A economia produz tudo o que é demandado	Há restrições de oferta
Métrica de produto	Ganho/perda monetário (econômico)	Ganho/perda de bem-estar
Acurácia das estimativas	Tende a superestimar impactos de políticas	Tende a subestimar impacto de políticas

Fonte: Adaptado a partir de [Koks et al \(2016\)](#).

Inicialmente, os modelos de Análise Insumo-Produto pressupõem tecnologia setorial que opera com base em coeficientes técnicos de produção fixos. Isso equivale a dizer que não há possibilidade de substituição de insumos no processo produtivo, ou que a função de produção é do tipo Leontief. Essa é uma hipótese condizente com o curto prazo, pois ela se adequa bem num ambiente de rigidez de preços e ausência de melhoria técnica. Os modelos CGE, por sua vez, admitem substituição de insumos e flexibilidade de preços. Eles incorporam em sua estrutura funções de produção do tipo Cobb-Douglas e de Elasticidade de Substituição Constante (CES), por exemplo. Assim sendo, esse arcabouço permite vislumbrar a evolução de variáveis endógenas do modelo, inclusive no longo prazo.

Note-se que a estrutura matemática por trás desses modelos difere bastante. Em geral, modelos IO são resolvidos com base em álgebra matricial simples. Essa simplicidade decorre, em parte, da linearidade inerente à tecnologia de coeficientes fixos de produção. Já os modelos CGE tem em seu bojo uma estrutura muito mais rebuscada. Há a possibilidade de se especificar funções (de produção ou de utilidade) não-lineares, embora seja recorrente se trabalhar também com funções agregativas lineares.²

Essas diferenças também repercutem diretamente no processo de interdependência setorial e na maneira como o efeito das políticas transborda e permeia a economia como um todo. Para a AIO, a interdependência setorial decorre dos coeficientes técnicos de produção, já nos modelos CGE essa interdependência é caracterizada através de elasticidades preço-cruzadas. Com efeito, enquanto na AIO a interdependência setorial vem de uma hipótese meramente tecnológica, nos modelos CGE, além da tecnologia, há uma hipótese calcada no processo de tomada de decisão dos agentes, que atentam para os preços e para a escassez dos insumos e bens e para a possibilidade de substituí-los, seja pela ótica do consumidor, seja pela ótica da firma.

A diferença nas estruturas teóricas acaba tendo reflexo direto sobre o efeito transbordamento dessas duas abordagens: enquanto a AIO é exclusivamente calcada em choques artificiais nos elementos que compõem a demanda, os modelos CGE geram transbordamento, respeitando tanto fatores de oferta (restrições de oferta, por exemplo) quanto de demanda.

² Do lado computacional, enquanto a AIO faz uso de softwares como Matlab, Octave, Scilab, e até mesmo Excel, os modelos CGE de maior porte já são resolvidos em softwares construídos exclusivamente para tal propósito, como o GAMS e o GEMPACK, por exemplo.

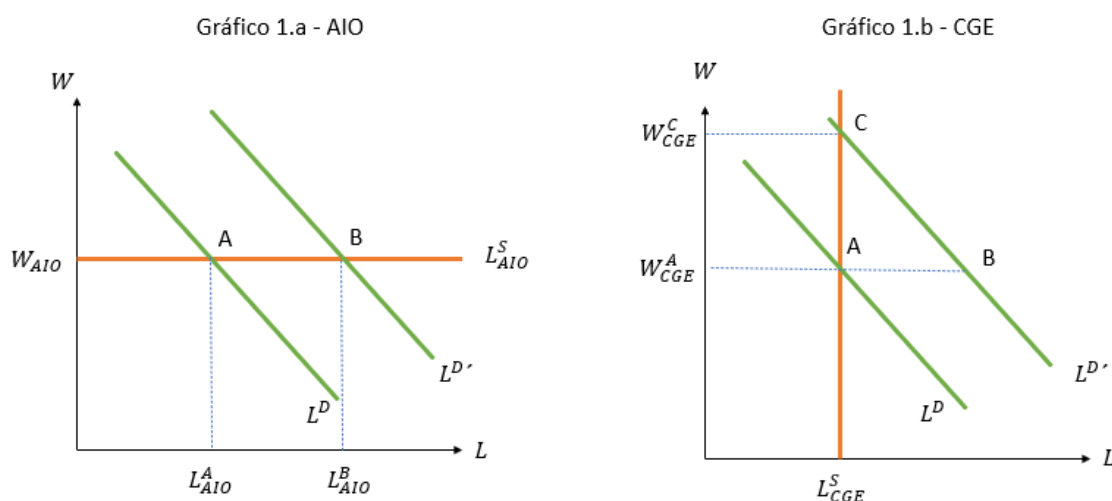
Por fim, há também diferenças nas métricas de produto usuais desses modelos: enquanto a AIO costuma discutir métricas monetárias puras, isto é, valores da produção em moeda corrente, os modelos CGE permitem uma análise com base em métricas teoricamente mais sofisticadas, como, por exemplo, níveis de utilidade e bem-estar econômico.

Em termos de avanços recentes, o que se observa é que ambos os modelos vêm expandindo sua complexidade, buscando introduzir e refinar elementos dinâmicos, assim como adequar-se às múltiplas regiões e à interdependência intra e inter-regional de uma melhor maneira.

Apesar dessa semelhança em termos de avanços, Gretton (2013) reforça seu ceticismo frente a modelagem AIO: ele argumenta que, nessa modelagem, i) faltam restrições do lado da oferta, ou seja, as curvas de oferta de fatores é perfeitamente elástica, ou ainda, toda a demanda pode ser prontamente atendida pelo lado da oferta; ii) os preços são fixos, o que não é muito condizente com uma análise de médio/longo prazo; iii) Há uma relação fixa para insumos intermediários à produção e resultados da produção; iv) os consumidores não responderam às mudanças de preço; e, v) O modelo pressupõe ausência de restrições orçamentárias domésticas e governamentais, ou seja, o modelo trata os novos gastos como uma espécie de “maná do céu”.

É possível contrastar a Análise Insumo Produto frente ao MARES/CE com base nas duas primeiras hipóteses levantadas anteriormente. Vejamos, por exemplo, quais as previsões dos dois modelos frente a um aumento na demanda por trabalho de um determinado setor. Como sustentam Fellows e Winter (2018), a oferta de trabalho é perfeitamente elástica para a AIO e os preços são fixos; já em modelos CGE, a oferta de trabalho é restrita à uma dotação fixa, o que sugere que uma elevação da demanda por mão-de-obra geraria um aumento no preço do trabalho (salário). O Gráfico 1, a seguir, mostra o efeito desse aumento na demanda por trabalho, que salta de L^D para $L^{D'}$, em cada um dos modelos:

GRÁFICO 1: CHOQUES DE DEMANDA PARA MODELAGENS AIO E CGE



O painel (a) retrata o caso AIO enquanto o (b) o caso CGE. No primeiro caso, não há escassez de trabalho e os ganhos econômicos vem em função de empregos gerados (o equilíbrio final retornaria uma quantidade de mão de obra empregada igual a L^B_{AIO} , superior a inicial, L^A_{AIO} , embora os salários fiquem mantidos em W_{AIO}). No caso do CGE, a elevação da demanda, dada a restrição na oferta de trabalho, não retorna uma quantidade superior de trabalhadores alocados no setor, mas, ainda assim, traz ganhos econômicos na forma de salários ($W^C_{CGE} > W^A_{CGE}$).

Outra questão que não é muito discutida na literatura é a questão da mudança tecnológica e estrutural. Em ambas as classes de modelos, alterações na tecnologia de um

determinado setor podem reverberar por toda a economia. É preciso, no entanto, buscar compreender ao certo as fontes dessa mudança tecnológica em cada uma dessas abordagens.

Para a AIO, a mudança estrutural equivale a uma alteração nos coeficientes diretos associados a inversa de Leontief, ou seja, essa mudança sugere uma alteração na maneira como se converte os insumos setoriais, que eram utilizados em uma proporção fixa e que passam a ser utilizadas numa nova proporção otimizada, em produto.

Já em modelos CGE essa mudança pode decorrer de uma alteração no fator de escala da função de produção de um bem composto que mixa capital e trabalho com base numa tecnologia Cobb-Douglas, por exemplo. Ou seja, a alteração pode ser decorrente de uma melhoria técnica que converte quantidades já ótimas de capital e trabalho em quantidades superiores de produto.

Discutiremos em maiores detalhes a questão da produtividade mais adiante, contudo, adentrar em cada uma das hipóteses dos modelos, como se esboçou acima, é algo que transpassa o esforço despendido nesse trabalho, assim como tornaria a leitura demasiadamente longa e entediante. Ainda assim, é importante observar que todos os pontos teóricos apontados trazem consigo reflexos possivelmente conflitantes para as predições dos modelos.

3 DA TEORIA PARA O EMPÍRICO

Em termos agregados, as Análises de Impacto Econômico no Estado do Ceará são realizadas basicamente através de dois modelos: Há as Análises de Insumo-Produto padrão e há o Modelo CGE do IPECE conhecido por MARES/CE (Modelo de Análises Regionais Estático para o Ceará). Ambas as abordagens partem do pressuposto de que se está numa posição de equilíbrio (geral) inicial e, após um determinado choque, convergem para uma posição de equilíbrio final.

Para se estabelecer o equilíbrio inicial, ambas as metodologias utilizam como fonte de dados informações disponíveis na Matriz Insumo-Produto (MIP) e na Tabela de Recursos e Usos (TRU). Para o Ceará, esses dois conjuntos de informações estão prontamente disponíveis, para o ano de 2013, no [site do IPECE](#). O MARES/CE faz uso dessas mesmas informações, embora elas sejam estruturadas numa Matriz de Contabilidade Social (SAM) e sejam necessárias algumas poucas informações adicionais. Em posse desses dados, calibram-se parâmetros para os modelos e computam-se valores para variáveis relevantes com base nos mesmos. Além disso, é possível se calcular uma série de indicadores.

Aqui se argumenta que é possível explorar pontos metodológicos relevantes para uma comparação entre as predições das duas abordagens. O primeiro ponto que exploramos refere-se ao potencial de encadeamento e à produtividade setorial.

3.1 Encadeamento vs Produtividade

Análises Insumo-Produto foram bastante utilizadas no Brasil para se justificar injeções de recursos em “setores-chaves” da economia. Abordaremos essa questão em maiores detalhes ao final do trabalho, mas, abstendo-se de qualquer controvérsia, os setores-chave seriam aqueles capazes de impulsionar a economia caso se expandissem, ou caso fossem incentivados.

A ideia de setores-chave é inerente à AIO e foi calcada nos estudos de [Rasmussen \(1956\)](#) e [Hirschman \(1958\)](#). Esses autores elaboraram dois índices que posteriormente tiveram algumas variações. De um modo geral, os “índices de ligações para trás” fornecem uma medida relativa de quanto, na margem, um determinado setor demandaria de insumos do seu setor e dos demais. Já os “índices de ligações para frente” sugerem uma medida relativa de quanto, também na margem, esse setor ofertaria de produtos para o seu setor e para os demais. Posto

isso, os setores-chaves da economia poderiam ser declarados como aqueles que detivessem fortes índices de ligação para frente e para trás.

Aqui cabe um adendo teórico: primeiramente, note-se que a teoria que orienta a AIO é toda ela dirigida pelo lado da demanda, e que não há uma função que mapeie o conceito de encadeamento à um conceito Neoclássico que naturalmente está calcado no lado da oferta. Ainda assim, seria interessante indagar se as modelagens AIO e CGE seriam capazes de indicar os mesmos “setores-chave” para a economia Cearense.

O MARES/CE tem em sua estrutura um alicerce Neoclássico. Não há uma definição Neoclássica precisa para “setores-chave”, entretanto, se imaginássemos que um determinado setor merecesse incentivos frente aos demais, então, em termos Neoclássicos, seria interessante que esse setor fosse o que apresentasse maior produtividade. Os exercícios conduzidos adiante admitem que os conceitos de encadeamento e produtividade são diretamente comparáveis. Essa é uma hipótese discutível e que motiva, inclusive, trabalhos futuros. De qualquer maneira, seguiremos aceitando tal hipótese como válida.

Definindo-se n como o número de setores em análise; ℓ_{ij} como sendo um elemento da matriz inversa de Leontief $L = (I - A)^{-1}$; L^* como sendo a média de todos os elementos de L , e; L_{*j} e L_{i*} como sendo respectivamente a soma de uma coluna e de uma linha típica de L . Então, o Índice de ligação para trás é definido por $U_j = [L_{*j} / n] / L^*$. Já o Índice de ligação para frente é definido por $U_i = [L_{i*} / n] / L^*$. Por definição, esses índices são normalizados, o que prescreve que valores maiores que a unidade indicariam setores acima da média, e, portanto, recomendariam forte encadeamento.³

Há diversos indicadores distintos de encadeamento para frente e para trás na literatura AIO. Aqui seguimos a ideia de Miller e Blair (2019) e buscamos formular um índice que capte tanto o encadeamento para frente quanto para trás. Para tanto, somamos os dois índices médios apresentados anteriormente, para cada setor, e ponderamos o valor resultante pelo maior valor setorial obtido. Esse é o nosso “indicador de encadeamento” apresentado na Tabela 2.⁴

Seguindo o raciocínio, precisamos agora de um indicador de produtividade para ser comparado ao indicador de encadeamento. No MARES/CE, os setores podem ser vistos como empresas representativas que maximizam lucro e, dessa forma, otimizam o emprego de fatores e o uso de insumos intermediários. Os insumos intermediários são utilizados em proporções fixas, juntamente com um bem composto que resulta da utilização de capital e trabalho. A demanda por esses dois fatores decorre do processo de maximização do lucro sujeito a restrição tecnológica existente. Mais especificamente, o problema da firma representativa do setor é:⁵

$$\max_{F_{h,i}} \pi_j^y = p_i^y Y_i - \sum_h p_h^f F_{h,i} \quad ; \quad Y_i = b_i \prod_h F_{h,i}^{\beta_{h,i}}$$

Onde: π_i^y : lucro do i -ésimo setor produtor do fator composto Y_i na primeira etapa (i ou j); Y_i : fator composto, produzido na primeira etapa e utilizado na segunda etapa pelo i -ésimo setor; $F_{h,i}$: o h -ésimo fator utilizado pelo i -ésimo setor na primeira etapa; p_i^y : preço do i -ésimo fator

³ Note-se que os índices de ligação para frente e para trás requerem apenas informações sobre a inversa de Leontief. Essa matriz já foi elaborada pelo IPECE e é disponibilizada na [Tabela09 do arquivo para download](#) disponibilizado no sitio do Instituto.

⁴ Em termos práticos, a ordenação setorial sugerida por esse índice diferiu muito pouco se comparada à do índice de ligação para frente médio. Os mais céticos podem confrontar esses índices com a análise de Campo de Influência apresentada mais adiante. Note-se também que, por ser um índice ponderado, esse índice não carrega consigo a ideia de que indicadores maiores que um retornam índices acima da média dos setores.

⁵ Ver Equação (2) de Paiva et al (2018) para maiores detalhes.

composto; p_h^f : preço do h -ésimo fator; $\beta_{h,i}$: coeficiente de participação na função de produção de fator composto; b_i : coeficiente de escala na função de produção de fator composto.

Note-se que o coeficiente b_i do modelo MARES/CE é nossa conhecida medida de produtividade da função de produção Cobb-Douglas. Assim, tomamos proveito dele enquanto indicador de produtividade para ser comparado ao indicador de encadeamento construído. Para efeito de comparação, esse indicador também foi ponderado pelo maior valor setorial calculado. A tabela a seguir apresenta o cômputo desses indicadores para os 16 macro setores para os quais se dispõe de dados.⁶

Diversos pontos surgem desse exercício. Primeiramente, notemos as convergências:

- i) O setor de “Serviços domésticos” tem a menor produtividade e encadeamento dentre os setores, seguido pelo setor de “Atividades Imobiliárias”;
- ii) “Administração, defesa, educação e saúde públicas e seguridade social” e “Agropecuária” também são tidos como setores de baixa produtividade/encadeamento: 13º e 14º lugar para AIO e 14º e 13º para CGE, respectivamente;
- iii) “Alojamento e alimentação” tem uma posição mediana, ficando na 8ª posição nos rankings de encadeamento e produtividade;
- iv) “Indústrias de Transformação” também tem uma posição equilibrada: esse setor fica em 7º no ranking AIO e em 6º no ranking CGE;
- v) Em melhor posição está o setor de “Comércio e reparação de veículos automotores e motocicletas”, que são ranqueados como o 4º setor de maior encadeamento/produtividade;
- vi) O setor de “Atividades profissionais, científicas e técnicas, administrativas e serviços complementares” tem excelente ranqueamento em ambos os casos: 1º lugar para AIO e 2º para CGE;

Há, entretanto, setores que revelam grande divergência entre encadeamento e produtividade:

- vii) O setor de “Construção” é o que chama mais atenção dos divergentes: enquanto ele é o de maior produtividade para a análise CGE, ele se configura apenas como o 10º setor em termos de encadeamento;
- viii) Já “Eletricidade e gás, água, esgoto, atividades de gestão de resíduos e descontaminação” é muito valorado pelo encadeamento (2º lugar), mas pouquíssimo valorado pela produtividade (11º lugar);
- ix) “Atividades financeiras, de seguros e serviços relacionados” também é relativamente bem valorado pelo encadeamento (5º lugar), mas pouco qualificado pela produtividade (12º lugar);

Por fim, há setores com uma divergência relativamente aceitável entre as abordagens AIO e CGE: Educação e saúde privada (12º e 10º); Indústria extrativa (11º e 9º); Serviços de artes, cultura, esporte, recreação e pessoais e organizações patronais, sindicais e outros serviços (9º e 7º) e “Informação e comunicação” (que ficam em 3º e 5º, respectivamente). Na próxima

⁶ A Análise Insumo produto requer que se trabalhe com matrizes quadradas para que se possa invertê-las. As únicas MIPs e TRUs quadradas que o IPECE disponibiliza são as de ordem 16x16, daí nossa opção por trabalhar com este número de setores.

subseção realizamos algumas simulações para tentar lançar luz sobre essas convergências e divergências.

3.2 – Simulações com Tecnologia Estática

Dada a divergência produtividade/encadeamento, seria interessante comparar os efeitos de um choque nos investimentos do setor de Construção. Como o investimento é um dos elementos da demanda final na equação de Leontief, e como o IPECE já disponibiliza a matriz inversa de Leontief, assim como o vetor de demanda final para cada atividade (penúltima coluna da Tabela03 da MIP 16x16), torna-se fácil calcular tal efeito para a AIO.

Seja L a inversa de Leontief e seja y o vetor de demanda final, então, o vetor de produto, x , necessário para atender essa demanda é dado por $x = Ly$. Nesses termos, um incremento de investimento de 1% no setor de construção geraria um novo vetor \tilde{y} , cuja produção requerida seria dada por: $\tilde{x} = L\tilde{y}$. Como os preços são fixos para a AIO, basta calcular a variação percentual entre estes valores de produto que se terá a taxa de crescimento do produto setorial. Para essa abordagem, o impacto da ampliação do investimento em 1% gera um crescimento do Produto Estadual de 0,09%. Dado que L não se altera, e que somente o elemento de \tilde{y} associado ao setor de Construção cresce 1%, é evidente que somente o elemento de \tilde{x} deste setor irá se ampliar em 1%, ou seja, apenas o produto setorial da construção deve crescer 1%.

Aqui cabem ressalvas: primeiro, e de modo mais preciso, o choque no investimento deveria ser dado na Formação Bruta de Capital Fixo do setor de Construção. Para este setor, a FBKF representa 100% da demanda final, ou seja, a ampliação do investimento é toda ela convertida para demanda final. Assim, basta que se amplie a demanda final para o setor de construção de 13.450 milhões para 13.585 milhões de reais.

TABELA 2 : INDICADOR DE ENCADEAMENTO (AIO) VS INDICADOR DE PRODUTIVIDADE (CGE)

Setores	Encadeamento (AIO)	Produtividade (CGE)	Ranking (AIO)	Ranking (CGE)
Agropecuária	0,699	0,781	14	13
Indústria extrativa	0,742	0,933	11	9
Indústrias de transformação	0,896	0,961	7	6
Eletricidade e gás, água, esgoto, atividades de gestão de resíduos e descontaminação	0,987	0,918	2	11
Construção	0,764	1,000	10	1
Comércio e reparação de veículos automotores e motocicletas	0,959	0,981	4	4
Transporte, armazenagem e correios	0,937	0,991	6	3
Alojamento e alimentação	0,792	0,945	8	8
Informação e comunicação	0,966	0,980	3	5
Atividades financeiras, de seguros e serviços relacionados	0,956	0,863	5	12
Atividades imobiliárias	0,697	0,537	15	15
Atividades profissionais, científicas e técnicas, administrativas e serviços complementares	1,000	0,998	1	2
Administração, defesa, educação e saúde públicas e seguridade social	0,711	0,686	13	14
Educação e saúde privada	0,711	0,933	12	10
Serviços de artes, cultura, esporte, recreação e pessoais e organizações patronais, sindicais e outros serviços	0,790	0,950	9	7
Serviços domésticos	0,609	0,503	16	16

Fonte: Elaboração dos autores

O segundo adendo é que nos referimos a valores estimados de produto. No modelo AIO, essa estimativa tem como base a matriz de Leontief e o vetor de Demanda Final (a preços básicos) provido pela MIP. Já para o MARES/CE definiremos como “produto” a variável Produção Estadual Bruta, dada por Z_i^S , a qual é descrita na expressão (8) do trabalho que apresenta o modelo ([link](#)). Por se tratar de estimativas para o PIB setorial, esses valores podem diferir um pouco. A Tabela 3, a seguir, apresenta os níveis (em milhares de R\$ a preços de 2013) e as participações relativas de produto setorial no produto estadual preconizadas por cada um desses modelos, ambos em equilíbrio inicial. Para efeito de comparação, também são apresentados os valores da oferta de produtos disponibilizada pelo IPECE na Tabela de Recursos e Usos (TRU).

TABELA 3 : PRODUTO (TRU) E PRODUTO ESTIMADO ATRAVÉS DE AIO E DO MARES/CE

Setor	TRU		AIO		MARES/CE	
	PIB	%	PIB	%	PIB	%
Agropecuária	6.289	3,81	5.402	3,44	6.639	4,03
Indústria extrativa	1.075	0,65	696	0,44	1.071	0,65
Indústrias de transformação	33.780	20,48	56.560	35,99	33.660	20,41
Eletricidade e gás, água, esgoto, atividades de gestão de resíduos e descontaminação	7.164	4,34	3.397	2,16	7.164	4,34
Construção	15.116	9,17	13.450	8,56	14.773	8,96
Comércio e reparação de veículos automotores e motocicletas	20.477	12,42	13.671	8,70	20.742	12,58
Transporte, armazenagem e correios	7.169	4,35	3.205	2,04	7.144	4,33
Alojamento e alimentação	5.929	3,60	4.790	3,05	5.950	3,61
Informação e comunicação	4.640	2,81	2.898	1,84	4.648	2,82
Atividades financeiras, de seguros e serviços relacionados	5.900	3,58	4.735	3,01	5.900	3,58
Atividades imobiliárias	9.977	6,05	8.463	5,39	9.714	5,89
Atividades profissionais, científicas e técnicas, administrativas e serviços complementares	10.113	6,13	3.170	2,02	9.401	5,70
Administração, defesa, educação e saúde públicas e seguridade social	27.684	16,79	27.684	17,62	28.556	17,32
Educação e saúde privada	4.785	2,90	4.519	2,88	4.698	2,85
Serviços de artes, cultura, esporte, recreação e pessoais e organizações patronais, sindicais e outros serviços	3.596	2,18	3.287	2,09	3.635	2,20
Serviços domésticos	1.222	0,74	1.222	0,78	1.222	0,74
Total	164.917	100	157.149	100	164.917	100

Fonte: Elaboração dos autores

Retornando ao exercício de comparação, para se operacionalizar esse mesmo choque no MARES/CE, pode se dar a mesma ampliação no investimento do setor de construção de 1%. Adicionalmente, é necessário se ampliar a poupança com um montante de recursos equivalente ao incremento do investimento. No exercício, optou-se por equiparar a elevação do investimento à poupança do resto do país, que é exógena. Os efeitos setoriais desse choque em termos de produto, descontada a variação nos preços, pode ser visto na Tabela 4, a seguir.

TABELA 4 : IMPACTO DO AUMENTO DE 1% NO INVESTIMENTO DO SETOR DE CONSTRUÇÃO NO PIB DO ESTADO – MARES/CE

Setores	Taxa de Crescimento (%)
Agropecuária	-0,414
Indústria extrativa	-26,270
Indústrias de transformação	0,568
Eletricidade e gás, água, esgoto, atividades de gestão de resíduos e descontaminação	0,199
Construção	0,953
Comércio e reparação de veículos automotores e motocicletas	-0,659
Transporte, armazenagem e correios	0,310
Alojamento e alimentação	1,797
Informação e comunicação	-0,763
Atividades financeiras, de seguros e serviços relacionados	-1,675
Atividades imobiliárias	0,290
Atividades profissionais, científicas e técnicas, administrativas e serviços complementares	-0,090
Administração, defesa, educação e saúde públicas e seguridade social	0,393
Educação e saúde privada	1,216
Serviços de artes, cultura, esporte, recreação e pessoais e organizações patronais, sindicais e outros serviços	1,136
Serviços domésticos	-0,147
Total	0,08

Fonte: Elaboração dos autores

Em conformidade com a literatura internacional, o modelo CGE apresenta impactos sobre o produto estadual e setorial inferiores ao da modelagem AIO. Na modelagem AIO o efeito da ampliação do investimento em 1% no setor de Construção gera uma ampliação de 1% no produto setorial e 0,09% no produto estadual. No MARES/CE, essas elevações são de, respectivamente, 0,953% e 0,08%.

Por que esse efeito é menor? Em parte, por conta das hipóteses de possibilidade de substituição e de restrição de recursos empregada nos CGEs. No MARES/CE, a ampliação do investimento no setor de construção gera uma realocação dos níveis de mão-de-obra e capital utilizados pelos setores. Os fatores capital e trabalho no setor de construção crescem 1,24% e 0,744%, respectivamente. De outro lado, há uma redução de capital e mão-de-obra na indústria extrativa (esses fatores têm queda de -26,184% e -26,462%, respectivamente), o que faz com que o produto nesse setor caia bastante (-26,270%). Note-se que as participações relativas desses dois setores no PIB também destoam bastante (a indústria extrativa representa apenas 0,65% do PIB, já o setor de construção, 9,17%), o que está em conformidade com a realocação setorial de fatores apresentada anteriormente. Ademais, a relação entre esses dois setores sugerida pelo modelo parece estar de acordo com a análise de [Trompieri Neto et al \(2018\)](#), que sustentam que ciclo do setor Extrativo de Minerais não apresenta integração com a dinâmica de curto prazo dos demais setores.

Também é interessante averiguar o impacto em alguns setores. É de se esperar, por exemplo, uma elevação do PIB setorial de Eletricidade e gás, água, esgoto, atividades de gestão de resíduos e descontaminação, assim como uma elevação do PIB setorial das atividades imobiliárias, o que de fato prescreve o modelo. Atividades financeiras, de seguros e serviços relacionados, entretanto, está negativamente relacionada com o investimento no setor de construção, indicando um resultado não esperado e que requer maiores investigações.

3.3 Campos de Influência e Contrafactual

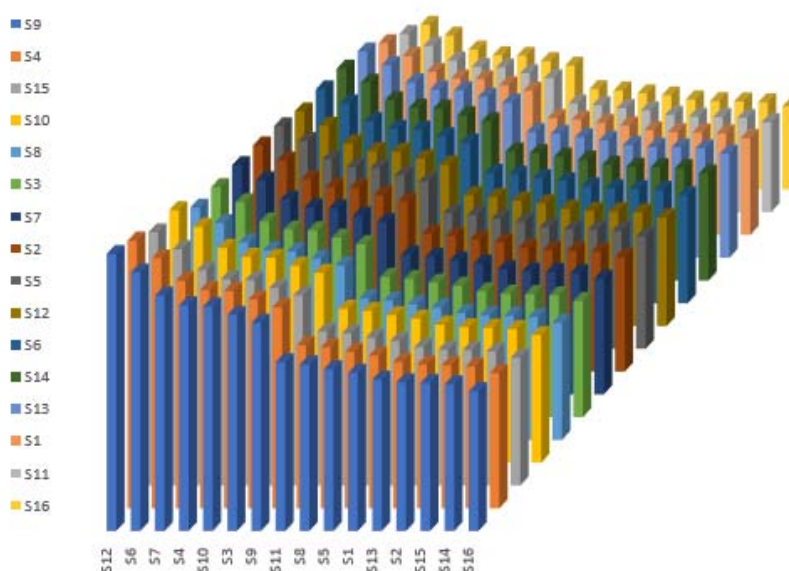
Uma das principais críticas dirigidas ao uso de modelos de insumo-produto foi sua incapacidade de lidar com as mudanças tecnológicas nos coeficientes de produção induzidas por novas inovações. Com o tempo, essas críticas foram sendo superadas através da introdução e aprimoramento de uma metodologia chamada de “campos de influência”. Essa abordagem passou a ser o principal veículo para descrever as mudanças gerais nas relações econômicas entre as indústrias causadas por combinações de mudanças nos coeficientes tecnológicos

advindas da difusão tecnológica e de inovações tecnológicas, organizacionais e administrativas [Sonis e Hewings (2009)]. Segundo os autores, a análise dos campos diretos de influência gera uma matriz “multiplicadora do produto”, e uma paisagem econômica artificial correspondente que representa a análise clássica dos setores-chave e as hierarquias das ligações setoriais para a frente e para trás.

O conceito de campo de influência mostra como se distribuem as mudanças dos coeficientes tecnológicos como um todo na economia, averiguando quais relações entre os setores são mais importantes dentro do processo produtivo. Além disso, é possível verificar as maiores influências de um dado setor sobre o restante da economia.

O Gráfico 2, a seguir, apresenta o computo do campo de influência para os 16 setores de atividades em análise. A ideia é que os setores dispostos no eixo das coordenadas seriam os que ofereceriam maior encadeamento para frente, ou seja, teriam forte potencial de ofertar produtos para os demais, já os setores dispostos nas colunas são os que mais apresentam encadeamento para trás, ou seja, que mais demandariam bens e serviços dos setores dispostos nas linhas.⁷

GRÁFICO 2 – CAMPO DE INFLUÊNCIA HIERARQUIZADO



Depreende-se da análise que inovações tecnológicas impulsionariam o crescimento através da provisão realizada pelos setores de Atividades profissionais, científicas e técnicas, administrativas e serviços complementares (S12), Comércio e reparação de veículos automotores e motocicletas (S6) e Transporte, armazenagem e correios (S7). Já os setores de Informação e comunicação (S9), Eletricidade e gás, água, esgoto, atividades de gestão de resíduos e descontaminação (S4) e de Serviços de artes, cultura, esporte, recreação e pessoais e organizações patronais, sindicais e outros serviços (S15) seriam setores de forte absorção.

Em termos práticos, a metodologia do campo de influência define um forte elo para a estrutura econômica do Ceará sustentado pela provisão do setor de Atividades profissionais, científicas e técnicas, administrativas e serviços complementares (S12), que seria absorvida pelo setor de Informação e comunicação (S9). É importante questionar se essa predição seria comportada pelo MARES/CE.

⁷ A Tabela A1, no apêndice, reporta os dados geradores do gráfico.

Conforme descreveu-se anteriormente, o MARES/CE admite em sua estrutura tanto insumos setoriais quanto bens de capital e trabalho para formulação da produção setorial, e esses fatores são combinados de maneira diferente da AIO.

No MARES/CE, o processo de melhoria técnica se dá através de uma melhoria exógena no produto por trabalhador. Para ver isso, tome uma função de produção Cobb-Douglas padrão: $Y = bK^\alpha L^{1-\alpha}$. Supondo retornos constantes de escala, então, tem-se a seguinte especificação de produto por trabalhador: $\frac{Y}{L} = b \left(\frac{K}{L}\right)^\alpha$. Ao supor uma unidade de capital por trabalhador, $\left(\frac{K}{L}\right) = 1$, então a produção gerada por essa unidade de capital por trabalhador será: $\frac{Y}{L} = b$, ou seja, o parâmetro b , que reflete a produtividade total dos fatores, nos revela qual o produto por trabalhador pode ser gerado por cada unidade de capital por trabalhador.

Posto isso, e contrastando essa ideia com as evidências do campo de influência apresentadas anteriormente, a pergunta que aqui se faz é: Quais seriam os efeitos (diretos e indiretos) no produto advindos dos setores de absorção (S9, S4 e S15) motivados por uma elevação da produtividade setorial dos setores de provisão (S12, S6 e S7)?

Se a predição da AIO estiver em linha com a do MARES/CE, então esse efeito deveria ser bastante “forte”. O MARES/CE permite que se investigue o espraiamento dos efeitos de um choque específico nos demais setores através de um processo de decomposição da taxa de crescimento da produção. Nós realizamos um exercício contrafactual dando um choque de 5% no parâmetro de produtividade setorial e observando de que modo as participações das taxas de crescimento das produções setoriais se comportavam frente a taxa de crescimento do produto Estadual. A Tabela 5, a seguir, reporta os resultados.

TABELA 5 : EFEITO DE CHOQUE NA PRODUTIVIDADE NO CRESCIMENTO DO PRODUTO (Setores Selecionados)

Setor		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	PIB
Comércio e reparação de veículos automotores e motocicletas	S6	-0,002	-0,058	0,477	0,036	-0,012	0,603	0,045	0,029	-0,004	-0,189	0,015	0,039	0,159	-0,043	0,007	-0,154	0,949
Atividades profissionais, científicas e técnicas, administrativas e serviços complementares	S12	0,005	-0,006	0,191	0,027	0,003	0,037	0,020	0,011	0,010	-0,057	0,008	0,083	0,069	0,008	0,008	-0,008	0,41
Transporte, armazenagem e correios	S7	0,004	0,006	0,142	0,012	-0,005	0,004	0,065	0,007	-0,006	-0,055	0,003	0,006	0,042	0,005	0,008	0,000	0,238
Média das Decomposições		0,003	-0,019	0,270	0,025	-0,004	0,215	0,044	0,016	0,000	-0,101	0,009	0,043	0,090	-0,010	0,008	-0,054	0,532
Participação Média do Impacto no PIB		0%	-4%	51%	5%	-1%	40%	8%	3%	0%	-19%	2%	8%	17%	-2%	1%	-10%	100%

Um choque de 5% na produtividade do setor de Atividades profissionais, científicas e técnicas, administrativas e serviços complementares (S12), por exemplo, geraria um crescimento do produto estadual da ordem de 0,41%, dos quais 0,027% seriam advindos do setor S4 (Eletricidade e gás, água, esgoto, atividades de gestão de resíduos e descontaminação), 0,010% do setor S9 (Informação e Comunicação) e 0,008% do setor S15 (Transporte, armazenagem e correios). Ou seja, a participação relativa do crescimento do produto desses setores no crescimento do PIB estadual, dado o choque de produtividade individual no setor S12, seria da ordem de apenas 7%, 2% e 2% para os setores S4, S9 e S15, respectivamente.

Ao se dar choques, também de 5%, no parâmetro de produtividade dos outros setores provedores observa-se, inclusive, valores negativos para o setor de Informação e Comunicação (S9). Em média, esses choques gerariam um crescimento do PIB estadual de 0,532%, e as participações setoriais relativas nesse crescimento seriam, em média, de apenas 5%, 0% e 1% para os setores S4, S9 e S15, respectivamente.

Esse efeito é bastante reduzido se comparado, por exemplo, com o efeito advindo dos setores da indústria de Transformação (S3) e de Comércio e reparação de veículos automotores e motocicletas (S6), que em média são da ordem de 51% e 40%. Fica evidente, então, que as predições desse exercício contrafactual com o MARES/CE não corroboram as predições do Campo de Influência.

A Tabela AII do apêndice apresenta esse exercício contrafactual de choque setorial de produtividade de 5% para todos os setores disponíveis. O efeito médio do ganho de produtividade face à Indústria de transformação é bastante elevado (64%) e vem acompanhado do setor da Administração, defesa, educação e saúde públicas e seguridade social (28%) e do setor de Comércio e reparação de veículos automotores e motocicletas (13%). Um setor que chama a atenção é o setor de Atividades financeiras, de seguros e serviços relacionados. O efeito médio do ganho de produtividade advindo desse setor é negativo e altivo (-13%).

4 CONCLUSÃO

Diversos pontos relevantes podem ser tirados da análise aqui conduzida. O primeiro é que, dadas as premissas teóricas de cada abordagem, era de se esperar que elas gerassem predições contraditórias, o que se confirmou.

Em nossa análise empírica, fizemos uma comparação entre as abordagens AIO e CGE para identificar os “setores-chave” da economia cearense. Essa comparação foi calcada na ideia de encadeamento-produtividade. Vimos que os setores com melhor encadeamento não são necessariamente os mais produtivos e vice versa. Simulamos um impacto de um incremento exógeno no investimento da ordem de 1% e vimos que a abordagem CGE previa impactos setoriais e agregados inferiores à AIO. Isso está em conformidade com a literatura que trata do tema e decorre, possivelmente, das hipóteses teóricas envolvidas em cada abordagem, como, por exemplo, possibilidade de substituição e rigidez de preços e de fatores.

Dentro da AIO, passamos para uma análise do campo de influência setorial, o qual sugeriu fortes elos entre os setores de Atividades profissionais, científicas e técnicas, administrativas e serviços complementares (S12), Comércio e reparação de veículos automotores e motocicletas (S6) e Transporte, armazenagem e correios (S7) - que seriam setores de provimento - com os setores de Informação e comunicação (S9), Eletricidade e gás, água, esgoto, atividades de gestão de resíduos e descontaminação (S4) e de Serviços de artes, cultura,

esporte, recreação e pessoais e organizações patronais, sindicais e outros serviços (S15), que seriam setores de absorção.

Conduzimos um experimento contrafactual através do modelo CGE onde simulamos ganhos de produtividade individuais para os setores de provimento indicados e constatamos que o efeito agregado dos setores de absorção sobre o PIB era relativamente baixo. Para todos os setores, em média, os setores que mais seriam capazes de impactar o produto por conta dos ganhos de produtividade setoriais seriam os setores da Indústria de transformação (S3) e da Administração, defesa, educação e saúde públicas e seguridade social (S13). Os resultados contrafactuais, portanto, também entram em conflito com os da análise do campo de influência.

Como lição, fica o fato de que se estabelecer “setores-chave” para a economia, mesmo com base em metodologias científicas consolidadas, exige um altíssimo grau de cautela. Metodologias distintas podem gerar resultados distintos e, até mesmo usando como base uma metodologia que se espera ser mais adequada, o subsídio ao setor de maior produtividade (Construção, conforme visto no caso do CGE para o Ceará) pode não ser a política direcionada que maximize produto.

Ademais, mesmo que se tenha fortes indícios agregados de que um determinado setor seja “chave”, a distribuição de firmas neste setor pode ter uma grande concentração de firmas ineficientes e pouquíssimas firmas eficientes. Neste caso, se um subsídio for dado ao setor, por exemplo, em termos de linhas de crédito, e se as empresas eficientes optarem por financiamento próprio enquanto as ineficientes tomarem os empréstimos, então os efeitos do incentivo seriam, na realidade, indesejáveis e perversos para a economia como um todo. Há toda uma literatura focada na ideia de “*misallocation*” - ou má alocação de recursos - que adentra nessa questão e que também carece ser mais bem discutida para o caso cearense em estudos futuros.⁸

Também é preciso ressaltar que não se está fazendo nenhum tipo de juízo de valor à cerca da AIO ou CGE. Todas as contribuições dessas linhas de pesquisa são bem-vindas e é possível se estender a análise em várias direções buscando um consenso ou uma corroboração das divergências. É possível se aprofundar a discussão teórica sobre o link produtividade-encadeamento. A AIO conduzida aqui também foi embasada em região única e é possível que matrizes de Leontief multi-regionais gerem resultados distintos. Um modelo CGE dinâmico também traria a possibilidade de novos resultados. É possível se converter o conceito de campo de influência para a abordagem CGE nos moldes do que fizeram Haddad e Hewings (2007), por exemplo. Todas essas questões apontam limitações do que aqui foi exposto e ficam como sugestão para estudos futuros.

⁸ Aos interessados sobre o tema, fica a sugestão de leitura de Restuccia e Rogerson (2017).

5 REFERÊNCIAS BIBLIOGRÁFICAS

- FIRME, Vinicius Azevedo Couto; PEROBELLI, Fernando Salgueiro. O setor energético brasileiro: uma análise via indicadores de insumo-produto e o modelo híbrido para os anos de 1997 e 2002. **Planejamento e políticas públicas**, n. 39, 2012.
- GRETTON, Paul. On input-output tables: uses and abuses. **Productivity Commission Staff Research Note**, 2013.
- HIRSCHMAN, Albert O. **The strategy of economic development**. 1958.
- HU, Aijun et al. Analyzing regional economic impact and resilience: a case study on electricity outages caused by the 2008 snowstorms in southern China. **Natural hazards**, v. 70, n. 2, p. 1019-1030, 2014.
- KOKS, Elco E. et al. Regional disaster impact analysis: comparing input–output and computable general equilibrium models. **Natural Hazards and Earth System Sciences**, v. 16, n. 8, 2016.
- MILLER, Ronald E.; BLAIR, Peter D. **Input-output analysis: foundations and extensions**. Cambridge university press, 2009.
- MONTOYA, Marco Antonio; PASQUAL, Cássia Aparecida. O uso setorial de energia renovável versus não renovável e as emissões de CO2 na economia brasileira: um modelo insumo-produto híbrido para 53 setores. 2015.
- OKUYAMA, Yasuhide. Globalization and localization of disaster impacts: an empirical examination. In: **CESifo Forum**. München: ifo Institut für Wirtschaftsforschung an der Universität München, 2010. p. 56-66.
- OKUYAMA, Yasuhide; SANTOS, Joost R. Disaster impact and input–output analysis. **Economic Systems Research**, v. 26, n. 1, p. 1-12, 2014.
- PAIVA, W. L.; TROMPIERI NETO, N. Revisitando os impactos econômicos do Acuario Ceará. IPECE Informe nº 83. Fortaleza: IPECE, 2015.
- RASMUSSEN, P. N. Studies in Intersectoral Relations, Amsterdam, North-Holland PC Schultz S.(1977) Approaches to Identifying Key Sectors Empirically by Means of Input-Output Analysis. **Journal of Development Studies**, v. 14, 1956.
- RESTUCCIA, Diego; ROGERSON, Richard. The causes and costs of misallocation. **Journal of Economic Perspectives**, v. 31, n. 3, p. 151-74, 2017.
- ROSE, Adam. Input-output economics and computable general equilibrium models. **Structural change and economic dynamics**, v. 6, n. 3, p. 295-304, 1995.
- SONIS, Michael; HEWINGS, Geoffrey JD. New developments in input-output analysis. In: **Tool Kits in Regional Science**. Springer, Berlin, Heidelberg, 2009. p. 69-117.
- TROMPIERI NETO, Nicolino et al.(2018) CARACTERIZAÇÃO DOS CICLOS DE NEGÓCIOS DOS SETORES PRODUTIVOS DA INDÚSTRIA BRASILEIRA. Working Paper CAEN/UFC
- WEST, Guy R. Comparison of input–output, input–output+ econometric and computable general equilibrium impact models at the regional level. **Economic Systems Research**, v. 7, n. 2, p. 209-227, 1995.

O EFEITO DA PANDEMIA DE COVID-19 NOS GASTOS PÚBLICOS MUNICIPAIS NO ESTADO DO CEARÁ: PROPOSIÇÃO E APLICAÇÃO DE MÉTODO DE DETECÇÃO DE *RED FLAGS*

Sarah Mesquita Lima*

Jaime de Jesus Filho**

Paulo Rogério Faustino Matos***

Ricardo Wagner Cavalcante Brito****

RESUMO

Motivada pelo período pandêmico pelo qual o mundo passa no ano de 2020, esta pesquisa objetivou propor e aplicar um método para detecção de red flags em despesas públicas municipais. Para tanto, após prévia busca na literatura, foram definidas métricas para aferição de padrões das despesas liquidadas pelo Poder Executivo, as quais foram aplicadas aos municípios cearenses com uso de dados informados pelos municípios ao Tribunal de Contas do Estado do Ceará. Com a utilização da técnica de boxplot, foram mapeados os municípios que apresentaram gastos anormais no período de março a julho de 2020 em função de sua população, utilizada como proxy para porte. A aplicação do método proposto permitiu identificar, considerando a classificação funcional da despesa, os municípios que demandam mais atenção, bem como quais as subfunções daqueles municípios devem ser examinadas com maior profundidade. O estudo traz contribuições aos campos acadêmico e profissional, pois ao primeiro acrescenta um conhecimento novo, representado uma combinação de métodos para solução de um problema; com relação ao segundo, disponibiliza aos agentes públicos, em especial aos analistas e auditores de Tribunais de Contas, uma metodologia que permite a detecção de red flags sem aplicação de técnicas de machine learning.

ABSTRACT

Motivated by the pandemic period the world is going through in the year 2020, this research aimed to propose and apply a method for detecting red flags in municipal public expenditure. To this end, after a previous search in the literature, it defined metrics to measure the patterns of expenses settled by the Executive Branch and applied them to the municipalities of Ceará using data reported by the municipalities to the State Audit Court. Using the boxplot technique, the municipalities that presented abnormal expenses in the period from March to July 2020 were mapped according to their population, used as a proxy for size. The application of the proposed method allowed to identify the municipalities that demand more attention. The research contributes with academic and professional aspects. Academically, as it adds new knowledge, represented by a combination of methods for solving a problem. Professionally, as it provides public agents with a methodology that allows the detection of red flags without applying machine learning techniques.

Área Temática: Setor público.

Classificação JEL: E62, G28, H72.

* Doutora em Administração de Empresas pela Unifor. Professora no IFCE. E-mail: sarah.mesquita@ifce.edu.br. Tel. (85) 9 9106-0275

** Doutor em Economia pela University of Chicago. Pesquisador no TCE-CE. E-mail: jjaimetilho@gmail.com.

*** Doutor em Economia pela FGV. Professor na UFC. E-mail: paulomatos@caen.ufc.br.

**** Doutorando em Ciências da Computação pela UFC. Professor na UniFB. E-mail: ricardowcb@gmail.com.

1 INTRODUÇÃO

Nos últimos dias do ano de 2019, na China, mais especificamente na cidade de Wuhan, foram registrados os primeiros casos de infecção pulmonar causada por um vírus até então desconhecido. Com o exponencial crescimento no número de casos e óbitos, bem como rápida difusão territorial da doença, a Organização Mundial da Saúde (OMS) declarou que o evento constituía uma Emergência de Saúde Pública de Importância Internacional (ESPII), em 30 de janeiro de 2020. E, em meados de março de 2020, a OMS declarou a pandemia de Covid-19.

No Brasil, em 3 de fevereiro de 2020, o Ministério da Saúde (MS), por meio da Portaria nº. 188/2020, declarou Emergência em Saúde Pública de Importância Nacional (ESPIN), o mais alto nível de alerta da Organização. No dia 6 de fevereiro de 2020, com objetivo de prover proteção à coletividade, estabelecendo as medidas para enfrentamento da Covid-19, foi sancionada a Lei nº. 13.979/2020. No entanto, o primeiro registro oficial no Brasil ocorreu em 26 de fevereiro de 2020.

No estado do Ceará, os primeiros casos foram oficialmente confirmados em 15 de março de 2020⁴ e, em 16 de março de 2020, por meio do Decreto nº. 33.510, foi decretada situação de emergência em saúde. Em virtude do crescente número de casos no estado, em especial na capital, os Decretos Legislativos nºs. 543 e 544 reconheceram estado de calamidade pública para o estado do Ceará e para a cidade de Fortaleza, sendo a calamidade do estado do Ceará reconhecida à nível federal em 28 de abril, oficializada por meio do Decreto nº. 33.555. Desde o início da pandemia até o final do mês de julho de 2020, 174 dos 184 municípios do estado do Ceará tiveram o estado de calamidade pública reconhecidos pela Assembleia Legislativa, o que representa quase 95% dos municípios cearenses⁵.

Importa destacar que, de acordo com o art. 65 da Lei Complementar nº. 101/2000 (Lei de Responsabilidade Fiscal – LRF), na ocorrência de calamidade pública reconhecida pelas Assembleias Legislativas, na hipótese dos Estados e Municípios, enquanto perdurar a situação, ocorre a flexibilização de algumas regras da LRF, dentre elas destacam-se (1) a suspensão da contagem dos prazos e as disposições estabelecidas nos arts. 23 e 70 (despesa total com pessoal) e 31 (dívida consolidada); e (2) a dispensa ao atingimento dos resultados fiscais e a limitação de empenho prevista no art. 9 (metas de resultado primário ou nominal).

Saliente-se ainda que a Lei nº. 8.666/93 (Lei de Licitações), no inciso IV de seu art. 24, estabelece que nos casos calamidade pública, a licitação de bens necessários ao atendimento da situação calamitosa e para as parcelas de obras e serviços que possam ser concluídas no prazo máximo de 180 (cento e oitenta) dias consecutivos e ininterruptos é dispensável a licitação. Saliente-se ainda que a, já citada, Lei nº. 13.979/2020 trouxe inovações importantes quanto à sistemática dos processos de licitações e formas de contratação pública.

Assim, o que deve gerar atenção, além do eventual descumprimento da meta fiscal, é a possibilidade da dispensa de procedimento licitatório para a aquisição de bens necessários ao atendimento da situação emergencial ou calamitosa, bem como a contratação de empresa que esteja com inidoneidade declarada ou com o direito de participar de licitação ou contratar com o Poder Público suspenso, quando se tratar da única fornecedora do bem ou serviço a ser adquirido. Referida dispensa pode facilitar irregularidades e culminar em fraude, superfaturamento, corrupção e/ou desvio de verba pública. De acordo com o ministro da

¹ OPAS Brasil (2020). Disponível [aqui](#).

² UNA-SUS (2020). Disponível [aqui](#).

³ SUS (2020). Disponível [aqui](#).

⁴ Sesa (2020). Disponível [aqui](#).

⁵ Diário do Nordeste (2020). Disponível [aqui](#).

Controladoria-Geral da União (CGU), Wagner Rosário, o governo tem enfrentado diversas tentativas de fraude nesse período pandêmico⁶.

Assim, esta pesquisa tem o objetivo de propor um método de detecção de *red flags* para auxiliar na identificação de indícios de fraude nos gastos públicos municipais com o uso de dados, bem como aplicá-lo.

2 PESQUISAS RELACIONADAS

Durante períodos de exceção, como o vivido em razão da pandemia de Covid-19, é natural que muitos entes decretam estado de calamidade pública, o que flexibiliza o atendimento a algumas regras da LRF. Em supracitado período, foi aprovada ainda a Lei n.º. 13.979/2020, a qual dispõe sobre as medidas para enfrentamento da ESPII decorrente do coronavírus responsável pelo surto de 2019 e trouxe inovações importantes quanto à sistemática dos processos de licitações e formas de contratação pública.

Assim, a combinação do previsto nas LRF e na Lei n.º. 13.979/2020 podem dar margem para aquisições e contratações fraudulentas, uma vez que, conforme Andvig (2008), estados de exceção (como é o caso da pandemia de Covid-19) podem potencializar fraudes.

Urge destacar que as despesas públicas comumente são vulneráveis à corrupção e à fraude. Niessen, Paciello and Fernandez (2020), citando o Escritório das Nações Unidas em Drogas e Crimes, indicam que o preço do público contratos é aumentado entre 10% e 25% por causa da corrupção globalmente. O Brasil, em 2019, obteve uma pontuação de 35 (de 100) no Índice de Percepção de Corrupção da Transparência Internacional Organização, o que o coloca em 106º em uma lista com 198 países⁷.

Destaque-se que a corrupção é definida como o abuso de um poder confiado a um indivíduo para obtenção de um retorno privado (Rose-Akerman & Palifka, 2016). Sua relação com outros fatores (econômicos, sociais, culturais etc.) vem sendo estudada pelos cientistas na busca de estabelecer suas causas ou principais determinantes, bem como identificar as consequências dos atos corruptos para a sociedade.

Nas últimas décadas, o número de trabalhos empíricos analisando a relação entre a corrupção e os mais diversos fatores cresceu, conforme explicitado em Lima (2017), as ênfases dadas pelos pesquisadores às pesquisas diferem, conforme sumarizado a seguir.

As pesquisas de Mauro (1995), Tanzi and Davoodi (1998), Silva, Freitas e Bandeira (2001) e Gupta, Davoodi and Alonso-Terme (2002) basearam-se em medidas de percepção sobre a corrupção, em virtude da dificuldade de mensuração da corrupção. Conforme sumarizado na Tabela 1.

⁶ Agência Senado Controladoria Geral da União (CGU). Disponível [aqui](#).

⁷ Transparency International (2020). Disponível [aqui](#).

Tabela 1 - Sumarização de pesquisas que analisaram a corrupção sob a perspectiva da percepção

Autoria	Resumo
Mauro (1995)	Analisa um conjunto de dados que consiste em índices subjetivos de corrupção, a quantidade de burocracia, a eficiência do sistema judicial e várias categorias de estabilidade política para uma série de países. Concluíram que a corrupção diminui o investimento, diminuindo assim o crescimento econômico.
Tanzi e Davoodi (1998)	Os pesquisadores, analisando um escândalo de corrupção ocorrido em Milão, observaram que após referido escândalo houve queda substancial nos gastos de capital. Assim, o artigo considera que não poderia ter sido uma coincidência, salientando que a corrupção está frequentemente ligada a projetos de capital.
Silva <i>et al.</i> (2001)	O artigo tem como objetivo analisar e medir a influência da corrupção na riqueza de um país. Admite-se implicitamente que o grau de desenvolvimento institucional tem um efeito adverso sobre a produtividade dos fatores de produção, o que implica em redução da renda per capita. Traz algumas evidências empíricas sobre os efeitos da corrupção na produtividade dos fatores, em uma amostra de 81 países estudados em 1998. A principal conclusão é que a corrupção afeta negativamente a riqueza de uma nação ao reduzir a produtividade do capital, ou sua eficácia.
Gupta <i>et al.</i> (2002)	A pesquisa discute vários canais pelos quais a corrupção pode afetar a desigualdade de renda e pobreza, concluindo que as políticas que reduzem a corrupção provavelmente reduzirão a desigualdade de renda e a pobreza.

Já, mais recentemente, as pesquisas de Auriol, Straub and Flochel (2016), Wang (2016), López-Iturriaga and Sanz (2018) e Niessen *et al.* (2020), com o uso de grandes volume de dados acerca das aquisições e/ou contratações públicas, desenvolveram soluções para detecção de indícios de fraudes a partir dos padrões identificados. Conforme sumarizado na Tabela 2.

Tabela 2 - Sumarização de pesquisas que analisaram indícios de corrupção com dados

Autoria	Resumo
Auriol <i>et al.</i> (2016)	A pesquisa apresenta evidências de fraude na contratação pública e seu impacto no desenvolvimento da economia, usando o método de predição linear. Utilizando os dados de quase 50.000 contratos públicos operações no Paraguai, de 2004 a 2007, os pesquisadores descobriram que no Paraguai a principal canal de corrupção é o uso de uma compra excepcional mecanismo, que é liberado das restrições exigido pelos outros mecanismos.
Wang (2016)	Empregando a teoria dos jogos, aprendizado de máquina e métodos estatísticos, detectou possíveis casos de fraude em contratos públicos.
López-Iturriaga and Sanz (2018)	Os pesquisadores, usando redes neurais, desenvolveram um sistema de alerta precoce de corrupção pública. A previsão do modelo desenvolvido é baseada na economia fatores e foi usado um banco de dados com os mais proeminentes casos de corrupção na Espanha.
Niessen <i>et al.</i> (2020)	Usando uma base de dados abertas, esta pesquisa desenvolveu modelo de aprendizagem não supervisionado para detecção de anomalias baseada no algoritmo de Isolation Forest. Os resultados mostram uma precisão de mais de 90% na detecção dessas anomalias conhecidas logo como na fase de licitação e na fase de contratação, indicando assim uma abordagem viável para detecção de anomalias nos contratos públicos.

Para auxiliar a identificação de possíveis fraudes, Magro e Cunha (2017) enfatizam a importância de a auditoria fazer uso de *red flags* (bandeira vermelha, em livre tradução). Para Dycus (2001), *red flags* representam faróis, fornecendo um aviso ao auditor, os quais são indicadores de que as circunstâncias ou condições presentes oferecem a oportunidade de cometer uma fraude.

Uma vez que uma bandeira vermelha é reconhecida, o auditor tem a responsabilidade de planejar procedimentos de auditoria para determinar se uma fraude foi cometida e, em caso afirmativo, em que extensão. Se um auditor não for treinado para reconhecer os sinais de alerta ou deixar de assumir a responsabilidade de auditar por fraude, a atividade fraudulenta provavelmente não será detectada (Dycus, 2001).

Conforme afirmado por Souza e Louzada (2020), algumas técnicas de detecção de indícios de fraude utilizam mecanismos elaborados e outras utilizam meios de detecção sem o emprego de sistemas computacionais baseados em aprendizado de máquina. Como exemplos de pesquisas realizadas com técnicas de identificação de *red flag* sem apoio de sistemas computacionais de aprendizado de máquina, podem ser citados Cuiabano, Leandro, Oliveira e Bogossian (2014), Tóth, Fazekas, Czibikn and Tóth (2014) e Tas (2017), os quais estão sumarizados na Tabela 3.

Tabela 3 - Sumarização de pesquisas que propuseram identificação de *red flags* sem aprendizado de máquina

Autoria	Resumo
Cuiabano <i>et al.</i> (2014)	A pesquisa objetiva revisar a literatura corrente sobre aplicação de filtros quantitativos (<i>screening</i>) para identificação de cartéis. Para a mensuram do risco de colusão de mercados ou setores específicos e detectam comportamentos suspeitos dos agentes desses mercados.
Tóth <i>et al.</i> (2014)	A pesquisa analisou e organizou o conluio sob a perspectiva de três dimensões, sendo elas: (1) meios de distorção da concorrência ou técnicas elementares de colusão; 2) partilha da renda; e 3) estrutura de mercado resultante.
Tas (2017)	A pesquisa, a partir de premissa de que não exige o histórico completo da licitação ou informações prévias detalhadas sobre a existência de atuação pretérita conjunta entre os potenciais licitantes, propõe uma metodologia para detecção de conluio. Partindo de argumentos teóricos sobre o comportamento na manipulação de ofertas, argumenta que o coeficiente do número de licitantes específico do leilão pode ser usado como uma medida de conluio, desenvolvendo equações para fazer o cálculo desse coeficiente.

Assim, diante de todo o apresentado, destaca-se o uso de *red flags*, que é um conceito bem difundido no âmbito das atividades de auditorias. Entende-se que a sua utilização pode auxiliar nas atividades dos auditores, bem como auxiliar no controle social dos gastos públicos.

Oliveira, De Caroli, Amaral e Vilca (2014), ao pesquisarem identificação de anomalias em registros contábeis, afirmam que as maiores anomalias não tenham sido identificadas pelo processo usual de amostragem estatística, salientando a relevância de técnicas de detecção de anomalias em grandes bases de dados.

A mineração de dados, de acordo com Fayyad, Piatetsky-Shapiro and Smyth (1996), é uma etapa do processo de *Knowledge Discovery in Database* (KDD), que consiste na aplicação da análise de dados e algoritmos de descoberta, os quais produzem uma enumeração particular de padrões ou modelos que visam o conhecimento sobre os dados. A Figura 1 evidencia uma visão geral das etapas que compõem o processo de KDD.

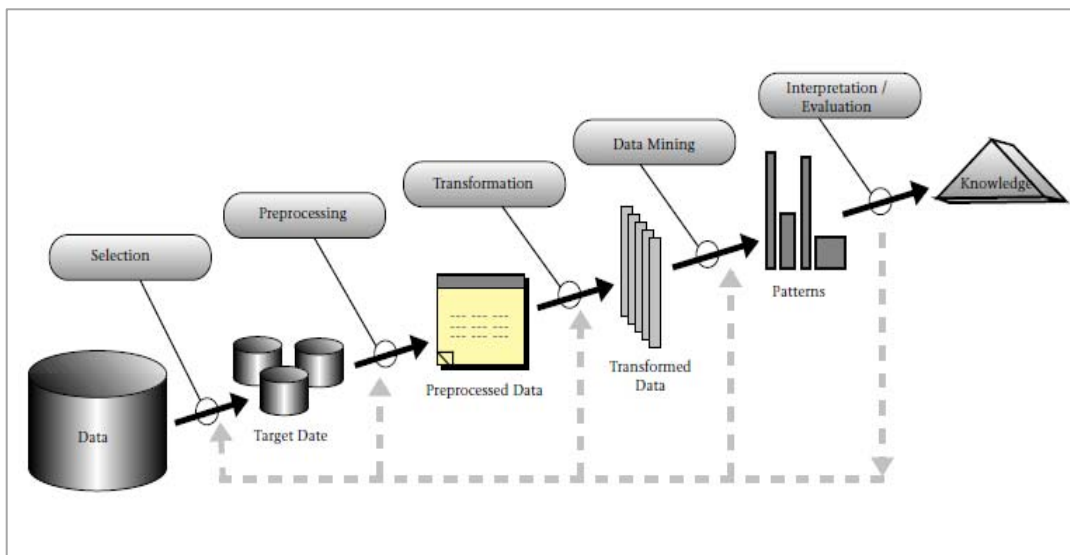


Figura 1: Etapas do processo KDD

Fonte: Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), p. 41.

Existem diversas formas de se analisar dados, então entender o objetivo que se espera alcançar com a mineração de dados é essencial. Assim, considerando o objetivo de auxiliar analistas e auditores na detecção de indícios de erros, fraudes ou anomalias nos gastos municipais, utilizará *boxplot* para detectar *red flags/outliers* nesta pesquisa, conforme evidenciado a seguir.

3 Metodologia

3.1 Técnica

Han, Pei and Kamber (2011), ao descreverem uma forma de análise de dados, orientam à suposição de que os dados para determinado atributo sejam classificados em ordem numérica crescente e que certos pontos de dados possam ser escolhidos de modo a dividir a distribuição de dados em conjuntos consecutivos de tamanhos iguais, conforme Figura 2.

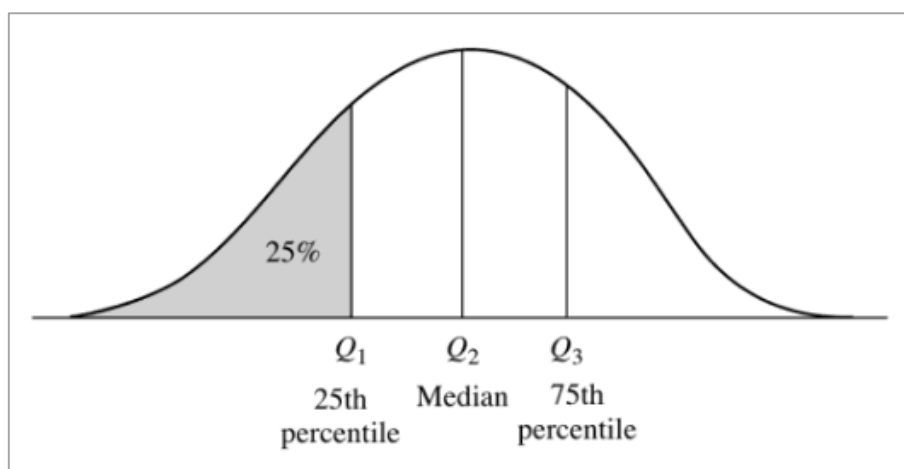


Figura 2: Representação da divisão de uma amostra em quartis

Fonte: Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier, p. 48.

Cada um desses pontos de divisão destacados na Figura 2 é denominado quartil. Os quartis, conforme evidenciado na Figura 2, fornecem uma indicação do centro, extensão e forma da distribuição. A partir da divisão de uma amostra em quartis, cinco atributos podem ser analisados, são eles:

- Menor valor (Min);
- Quartil inferior (Q1);
- Mediana (Q2);
- Quartil superior (Q3); e
- Maior valor (Max).

Saliente-se ainda um importante indicador de análise: o intervalo interquartil, o qual é reconhecido pela sigla IQR. O IQR é bastante utilizado na identificação de *outliers* e é definido como

$$\text{IQR} = \text{Q3} - \text{Q1} \quad (1)$$

Considera-se anomalia, os pontos situados além de $1,5 \cdot \text{IQR}$ inferior a Q1 ou $1,5 \cdot \text{IQR}$ superior a Q3 (Hawkins, 1980; Han *et al.*, 2011). Uma popular forma de visualizar a distribuição de uma amostra – e, partir de então, identificar *red flags/outliers* – é a utilização de *boxplots*, conforme evidenciado na Figura 3.

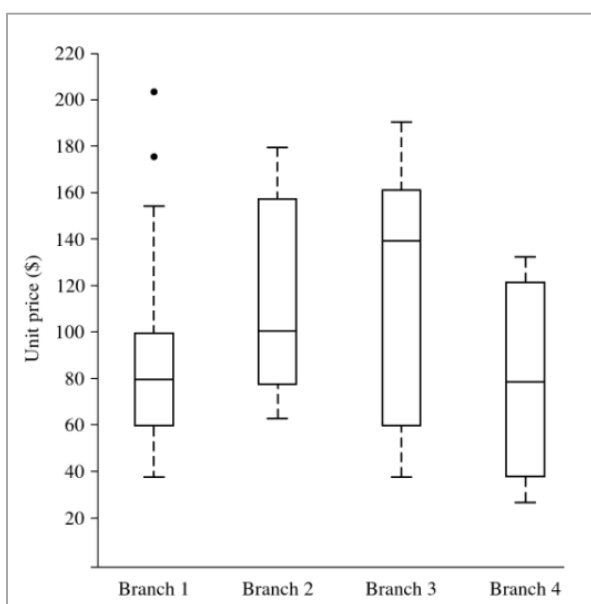


Figura 3: Representação de *boxplots*

Fonte: Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier, p. 51.

Utilizando o exemplo de Han *et al.* (2011), evidenciado na Figura 3, percebe-se que o preço unitário na Filial 1 (*Branch 1*) varia entre, aproximadamente, 40 e 200, sendo a mediana (Q2), Q1 e Q3, 80, 60 e 100 respectivamente. Assim, o IQR é 40 (Q3-Q1). Considerando o anteriormente descrito, podem ser consideradas anomalias valores abaixo de zero ($Q1 - 1,5 \cdot \text{IQR}$) e superior a 160 ($Q3 + 1,5 \cdot \text{IQR}$), que, na Figura 3, estão representados pelos pontos após o *whisher* superior.

3.2 Dados

Para realização das análises, foram utilizados dados oriundos do Sistema de Informações Municipais (SIM) do Tribunal de Contas do Estado do Ceará (TCE-CE), o qual, de acordo com o site do TCE-CE⁸, consiste, essencialmente, na definição de um padrão e orientações referentes à implantação, padronização, configuração, modulação, formatação dos dados, para que as informações das prestações de contas mensais das administrações municipais sejam registradas em meio informatizado e enviadas ao TCE-CE.

Para tanto, foram reunidas as despesas liquidadas realizadas entre os meses de março e julho entre os anos de 2017 e 2020. O estágio de liquidação da despesa foi eleito como o mais adequado por, de acordo com o Art. 63 da Lei n.º 4.320/1964, consistir na verificação do direito adquirido pelo credor. Por sua vez, o período selecionado se deu em razão de março ser o início do período pandêmico, conforme descrito na introdução; julho em razão da periodicidade de disponibilização de dados dos municípios ao TCE-CE; e 2017 a 2020 por se referirem ao mesmo ciclo de gestão municipal.

Saliente-se que foram analisadas as despesas informadas de qualquer natureza, com exceção daquelas relacionadas à Câmara Municipal, por estas representarem outra esfera de Poder e, por sua natureza, não atuar em ações diretamente relacionadas ao combate à pandemia. Por fim, destaque-se que, além dos dados oriundos do TCE-CE, foram utilizados dados do Instituto Brasileiro de Geografia e Estatística (IBGE).

3.3 Método proposto

Deste modo, no sentido de identificar *red flags* relacionadas às despesas liquidadas, foi proposto um processo de averiguação auxiliar, conforme representado na Figura 4.

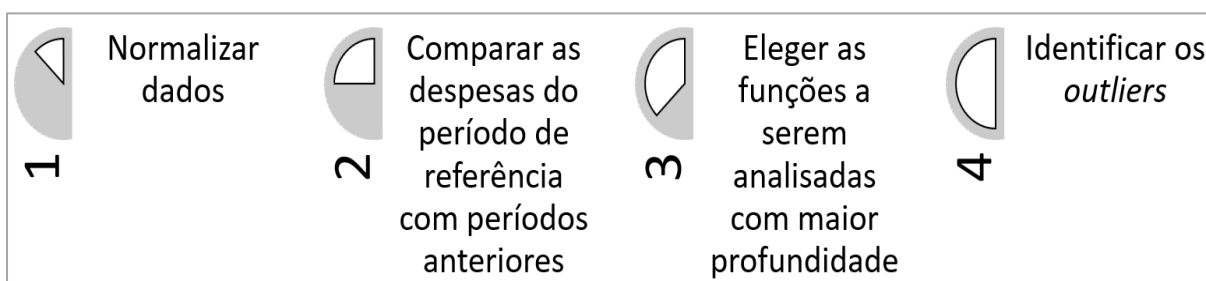


Figura 4: Estágios para identificação de *red flags*.

O estágio 1 prevê a normalização dos dados analisados. Na aplicação aqui proposta, referida normalização será feita em função da população municipal estimada pelo IBGE, como forma de neutralizar possíveis vieses decorrentes do porte municipal no valor das despesas liquidadas pelos municípios.

Em relação ao estágio 2, na aplicação, a comparação do período analisado (março a julho) de 2020 com o mesmo período dos anos anteriores da gestão municipal, será segregada em razão das despesas liquidadas por função, segundo a classificação funcional da Portaria n.º 42/1999, sendo levado em consideração a atualização pelo Índice de Preços para o Consumidor

⁸ TCE-CE (2020). Disponível [aqui](#).

Ampla (IPCA). Elegeram-se fazer referida segregação, pois as diferentes funções de despesa atendem a perfis diferentes de demandas, as quais apresentam padrões e ordenadores distintos.

A partir da comparação anteriormente descrita, de acordo com o previsto no estágio 3 da Figura 4, devem ser eleitas para análises com maior grau de precisão as funções que evidenciarem crescimento na liquidação das despesas acima do IPCA ou as funções que concentrem grande volume de recursos liquidados.

Por fim, para identificação e análise de *outliers*, conforme estágio 4 da Figura 4, uma sequência de análises foi proposta, é ela: (1) análise visual da relação entre as despesas liquidadas na função e a população; (2) aplicação da regressão linear para aferição estatística da relação entre as despesas liquidadas na função e a população; (3) identificação dos municípios que apresentam gasto/habitante como *outlier*; (4) análise das subfunções que apresentaram maior variação de 2019 para 2020 dos municípios com gasto/habitante como *outlier*.

Em relação à análise visual da relação entre as despesas liquidadas na função e a população, esta será realizada por meio de Gráfico de Dispersão. Saliente-se que, para melhor visualização dos dados ambos os eixos – despesas liquidadas e população – serão apresentados em escala logarítmica.

Para aferição da significância da relação entre as despesas liquidadas na função e a população, serão apuradas, com auxílio de linguagem de programação Python, conforme Apêndice 1, as regressões lineares, as quais têm como variável independente o número de habitante e como variável dependente o valor liquidado. A escolha do número de habitantes como determinante se deu por estar sendo considerado todo o universo de despesas municipais, as quais possuem diversas especificidades.

Com a verificação da relação existente entre as variáveis analisadas nas funções eleitas, serão, finalmente, elaborados *boxplots* para aferição de *outliers* em relação à variável “logaritmo das despesas liquidadas dividido pelo logaritmo da população”, seguindo as regras apontadas no anteriormente. Sendo considerado *outliers* os municípios que apresentarem-se com o valor da variável analisada com valores acima de $1,5 \cdot \text{IQR}$, seguindo o estabelecido por Hawkins (1980) e referenciado por Han *et al.* (2011). Por fim, propõe-se que sejam analisadas as variações das subfunções nas funções dos municípios que apresentaram-se como *outliers* na análise dos *boxplots*.

4 APLICAÇÃO DO MÉTODO

Tendo sido realizado os estágios 1, 2 e 3 anteriormente descritos, elegeram-se algumas funções para análises mais aprofundadas, seja pelo crescimento no volume de recursos liquidados, seja pela alta representatividade de referidas despesas no valor total liquidado, foram elas: Administração (função 04), Assistência Social (função 08), Saúde (função 10), Trabalho (função 11), Educação (função 12), Urbanismo (função 15), Gestão Ambiental (função 18), Comunicações (função 24), Energia (função 25) e Transporte (função 26).

Assim, serão observadas as despesas liquidadas relativamente ao número de habitantes⁹ com o objetivo de identificar *outliers*, os quais são importantes norteadores na detecção indícios de fraude e/ou corrupção. Considerando o objetivo de analisar as despesas municipais neste período pandêmico, bem como a limitação de páginas para escrita das análises, nas subseções a seguir, serão aplicados o estágio 4 do método proposto às despesas liquidadas

⁹ Eleita como *proxy* para porte dos municípios

de março a julho de 2020 em 3¹⁰ das 10 funções eleitas no estágio 3, são elas: Assistência Social (função 08), Saúde (função 10) e Gestão Ambiental (18).

4.1 Aplicação do método - Assistência Social

Com vistas a analisar com mais profundidade as despesas relacionadas à função Assistência Social, considera-se importante observar o Diagrama de Dispersão das variáveis valor liquidado na função Assistência Social e população estimada evidenciados em escala logarítmica, conforme Figura 5.

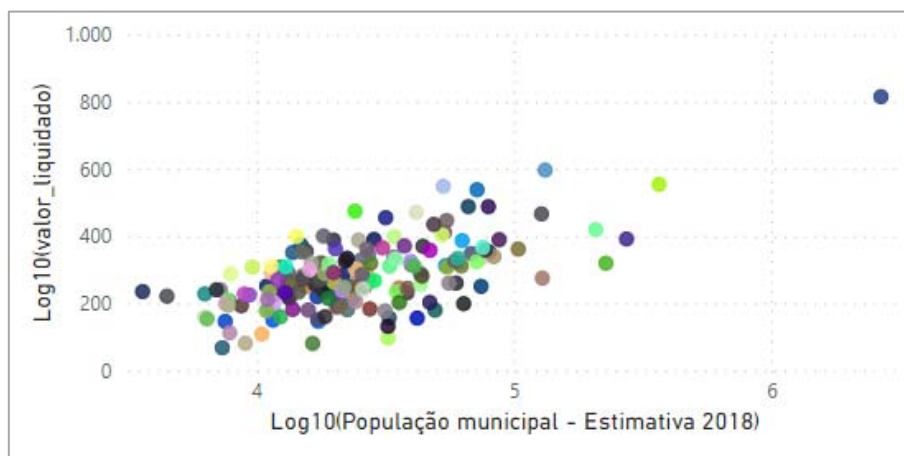


Figura 5: Dispersão do logaritmo das despesas liquidadas na função Assistência Social de março a julho do ano de 2020 e do logaritmo da população estimada de 2018

A Figura 5 evidencia que as despesas liquidadas na função Assistência Social de março a julho de 2020 parecem correlacionar-se com as populações municipais. Assim, para descrever o relacionamento entre as variáveis, foi calculada a regressão linear simples¹¹ tendo como variável independente o valor liquidado de março a julho de 2020 e como variável dependente a população estimada, a qual evidenciou os resultados evidenciados na Tabela 4.

Tabela 4 - Resultado da regressão linear simples das variáveis valor liquidado¹ na função Assistência Social e população estimada

Parâmetro	Valor
<i>Slope</i>	0,47
<i>Intercept</i>	1,63
<i>r</i>	0,25
<i>p-value</i>	1,79 ^{e-209}

¹De março a julho de 2020.

Os resultados evidenciados na Tabela 4 confirmam a correlação existente entre as variáveis, a qual, embora baixa, revela a significância estatística da relação, sendo referida relação descrita pela equação abaixo evidenciada:

$$Y = 1,63 + 0,47x \quad (2)$$

Onde: *Y* = Despesa liquidada na função Assistência Social; *X* = População estimada.

Assim, com o objetivo de identificar *outliers*, foram analisados *boxplots* das despesas liquidadas na função Assistência Social, conforme Figura 6.

¹⁰ Escolhidas pela robustez do incremento nas despesas no período analisado em 2020.

¹¹ Utilizando a linguagem de programação Python, conforme Apêndice 1.

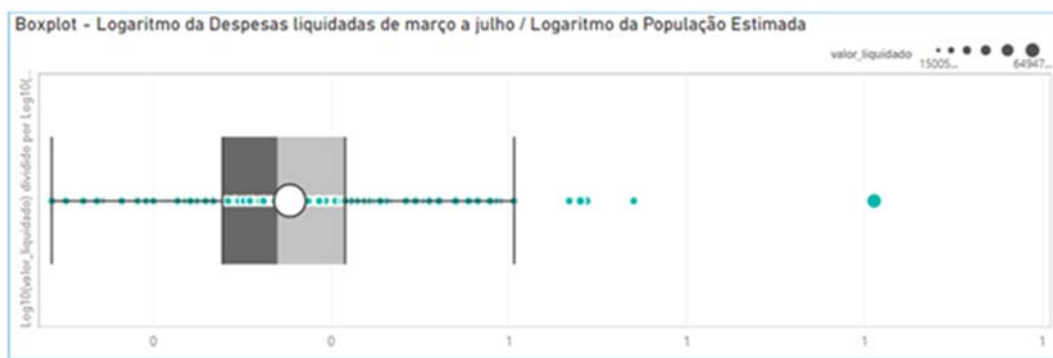


Figura 6: **Boxplot** da razão entre os logaritmos de despesas liquidadas na função Assistência Social e população para todos os municípios do estado do Ceará.

Os municípios (representados pelos pontos) que aparecem após os whisker (bigodes), conforme já explicitado, são dos que merecem maior atenção, uma vez que estão no quartil superior, especialmente os mais à direita da Figura 6, uma vez que representaram maior gasto por habitante na função analisada. Assim, o resultado dos boxplots, conforme Figura 6, evidencia que 5 dos 184 municípios como outliers, são eles: Fortaleza, Crato, Caucaia, Eusébio e Cascavel, nesta ordem. Dos 5 municípios, 4 tiveram seus estados de calamidade pública reconhecidos pela Assembleia Legislativa do Estado do Ceará.

Para uma análise mais aprofundada dos municípios acima elencados, recomenda-se que seja feita uma análise da variação dos valores liquidados entre os anos de 2019 e 2020 na função Assistência Social por subfunção em referidos municípios. A seguir, a título ilustrativo, será apresentada, na Figuras 7, supracitada análise para o município de Fortaleza.



Figura 7: **Gráfico de cascata** das despesas liquidadas de março a julho dos anos de 2019 e 2020 na função Assistência Social por subfunção – Município de Fortaleza

A análise da Figura 7, que evidencia a variação de 2019 a 2020 das despesas liquidadas pelo poder executivo do município de Fortaleza na função Assistência Social por subfunção, permite constatar que as maiores variações em valores absolutos estão associadas a acréscimos na subfunção Assistência Comunitária (244) seguida da subfunção Assistência à Criança e ao Adolescente (243). No entanto, em valores relativos, importa destacar o acréscimo de mais de

84% das despesas liquidadas na subfunção Assistência Comunitária (244) e o decréscimo de mais de 48% na subfunção Alimentação e Nutrição (306).

4.2 Aplicação do método - Saúde

Para analisar com mais profundidade as despesas relacionadas à função Saúde, seguindo as ações propostas, considera-se importante observar o Diagrama de Dispersão das variáveis valor liquidado em referida função e população estimada evidenciados em escala logarítmica, conforme Figura 8.

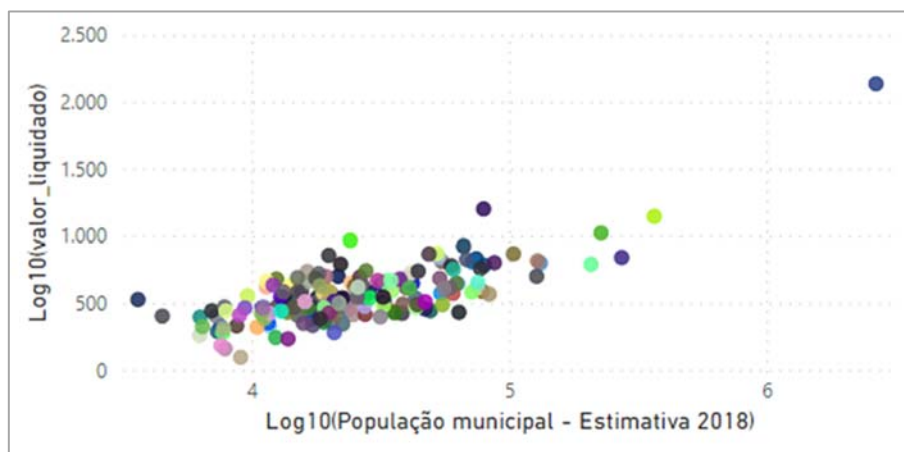


Figura 8: Dispersão do logaritmo das despesas liquidadas na função Saúde de março a julho do ano de 2020 e do logaritmo da população estimada de 2018

A Figura 8 evidencia que as despesas liquidadas na função Saúde de março a julho de 2020 parecem correlacionar-se com as populações municipais. Assim, para descrever o relacionamento entre as variáveis, foi calculada a regressão linear simples¹² tendo como variável independente o valor liquidado de março a julho de 2020 e como variável dependente a população estimada, a qual evidenciou os resultados evidenciados na Tabela 5.

Tabela 5 - Resultado da regressão linear simples das variáveis valor liquidado¹ na função Saúde e população estimada

Parâmetro	Valor
<i>Slope</i>	0,50
<i>Intercept</i>	2,02
<i>r</i>	0,25
<i>p-value</i>	0,00

¹De março a julho de 2020.

Os resultados evidenciados na Tabela 5 confirmam a correlação existente entre as variáveis, a qual, embora baixa, revela a significância estatística da relação, sendo referida relação descrita pela equação abaixo evidenciada:

$$Y = 2,02 + 0,5x \quad (3)$$

Onde: Y = Despesa liquidada na função Saúde; X = População estimada.

Assim, com o objetivo de identificar *outliers*, foram analisados *boxplots* das despesas liquidadas na função Saúde, conforme Figura 9.

¹² Utilizando a linguagem de programação Python, conforme Apêndice 1.

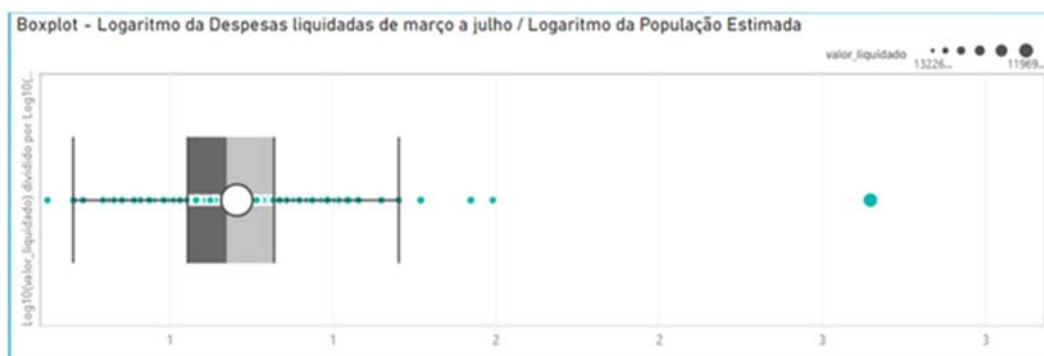


Figura 9: **Boxplot da razão entre os logaritmos de despesas liquidadas na função Saúde e população para todos os municípios do estado do Ceará.**

O resultado do *boxplot*, conforme Figura 9, evidenciam que 4 dos 184 municípios como outliers, são eles: Fortaleza, Quixeramobim, Caucaia e Maracanaú, nesta ordem. Dos 4 municípios, 3 tiveram seus estados de calamidade pública reconhecidos pela Assembleia Legislativa do Estado do Ceará. Dos 4 municípios, 3 tiveram seus estados de calamidade pública reconhecidos pela Assembleia Legislativa do Estado do Ceará.

Para uma análise mais aprofundada dos municípios acima elencados, recomenda-se que seja feita uma análise da variação dos valores liquidados entre os anos de 2019 e 2020 na função Saúde por subfunção em referidos municípios. A seguir, a título ilustrativo, será apresentada, na Figuras 10, supracitada análise para o município de Quixeramobim.

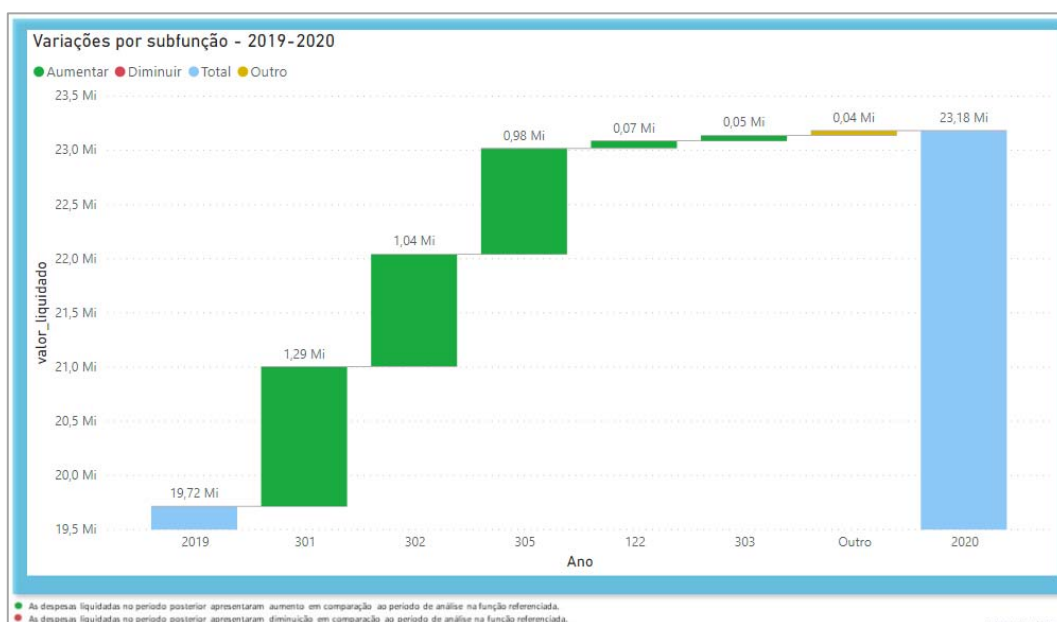


Figura 10: **Gráfico de cascata das despesas liquidadas de março a julho dos anos de 2019 e 2020 na função Saúde por subfunção – Município de Quixeramobim**

A análise da Figura 10, que evidencia a variação de 2019 a 2020 das despesas liquidadas pelo poder executivo do município de Quixeramobim na função Saúde por subfunção, permite constatar que houve um acréscimo dos valores liquidados em 2020 em comparação à 2019. Em valores absolutos cabe destacar as maiores variações foram acréscimos nas subfunções Atenção Básica (301) e Assistência hospitalar e ambulatorial (302). Em termos relativos, destaque-se um aumento de 185,2% na subfunção Vigilância Epidemiológica (305) e de 24,1% na subfunção Atenção Básica (301).

4.3 Aplicação do método – Gestão Ambiental

Para analisar com mais profundidade as despesas relacionadas à função Gestão Ambiental, seguindo as ações propostas, considera-se importante observar o Diagrama de Dispersão das variáveis valor liquidado em referida função e população estimada evidenciados em escala logarítmica, conforme Figura 11.

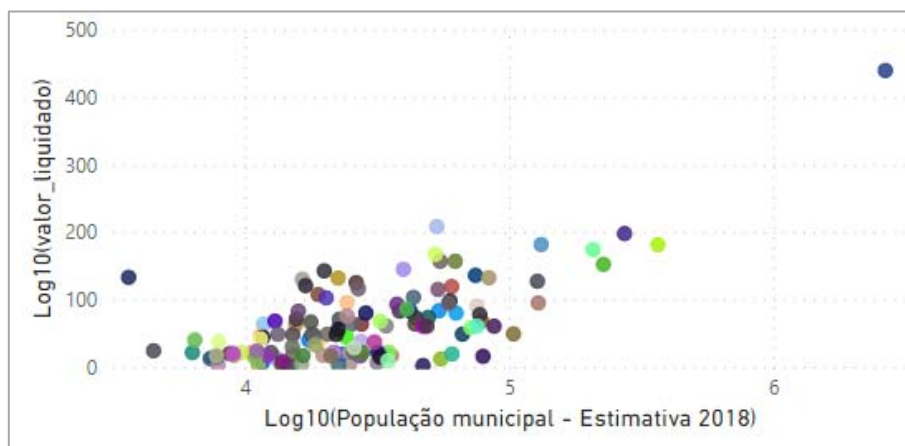


Figura 11: **Dispersão do logaritmo das despesas liquidadas na função Gestão Ambiental de março a julho do ano de 2020 e do logaritmo da população estimada de 2018**

A Figura 11 evidencia que as despesas liquidadas na função Gestão Ambiental de março a julho de 2020 parecem correlacionar-se com as populações municipais. Assim, para descrever o relacionamento entre as variáveis, foi calculada a regressão linear simples tendo como variável independente o valor liquidado de março a julho de 2020 e como variável dependente a população estimada, a qual evidenciou os resultados evidenciados na Tabela 6.

Tabela 6: **Resultado da regressão linear simples das variáveis valor liquidado¹ na função Gestão Ambiental e população estimada**

Parâmetro	Valor
<i>Slope</i>	0,39
<i>Intercept</i>	1,93
<i>r</i>	0,24
<i>p-value</i>	8,22 ^{e-31}

¹De março a julho de 2020.

Os resultados evidenciados na Tabela 6 confirmam a correlação existente entre as variáveis, a qual, embora baixa, revela a significância estatística da relação, sendo referida relação descrita pela equação abaixo evidenciada:

$$Y = 1,93 + 0,39x \quad (4)$$

Onde: Y = Despesa liquidada na função Gestão Ambiental; X = População estimada.

Assim, com o objetivo de identificar *outliers*, foram analisados *boxplots* das despesas liquidadas na função Gestão Ambiental, conforme Figura 12.

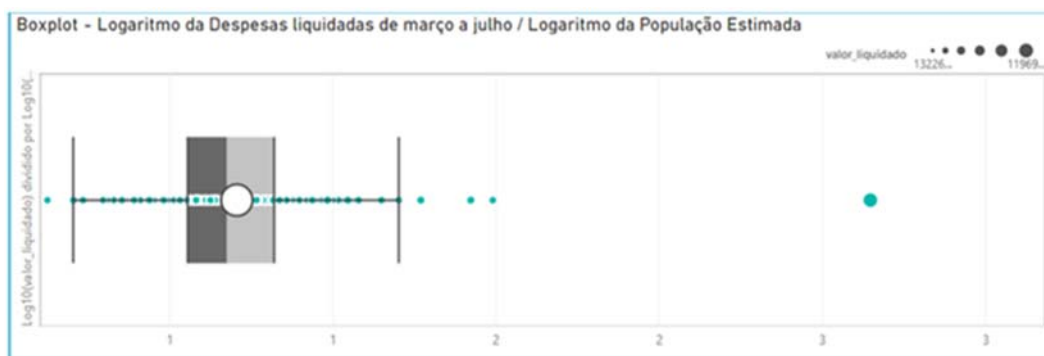


Figura 12: **Boxplot da razão entre os logaritmos de despesas liquidadas na função Gestão Ambiental e população para todos os municípios do estado do Ceará.**

Ao todo, 135 dos 184 municípios liquidaram despesas na função Gestão Ambiental. Assim, o resultado do *boxplot*, conforme Figura 12, evidencia que 4 dos 135 municípios como *outliers*, são eles: Fortaleza, Eusébio, Juazeiro do Norte e Crato, nesta ordem, tendo todos seus estados de calamidade pública reconhecidos pela Assembleia Legislativa do Estado do Ceará.

Para uma análise mais aprofundada dos municípios acima elencados, recomenda-se que seja feita uma análise da variação dos valores liquidados entre os anos de 2019 e 2020 na função Gestão Ambiental por subfunção em referidos municípios. A seguir, a título ilustrativo, será apresentada, na Figuras 13, supracitada análise para o município de Eusébio.

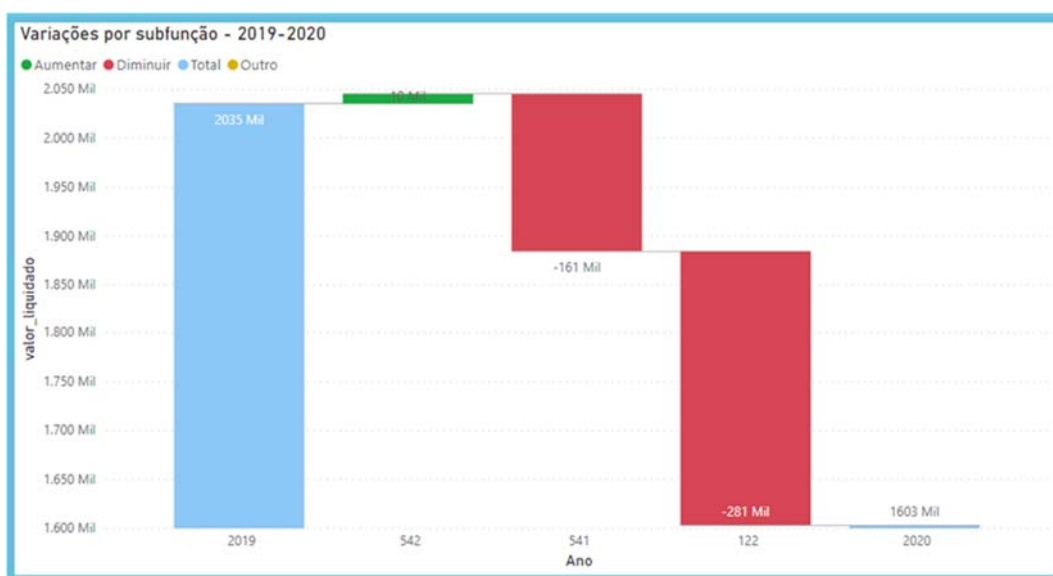


Figura 13: **Gráfico de cascata das despesas liquidadas de março a julho dos anos de 2019 e 2020 na função Gestão Ambiental por subfunção – Município de Eusébio**

A análise da Figura 13, que evidencia a variação de 2019 a 2020 das despesas liquidadas pelo poder executivo do município de Eusébio na função Gestão Ambiental por subfunção, permite constatar que houve um decréscimo dos valores liquidados em 2020 em comparação à 2019, ou seja, o período pandêmico não ocasionou elevação em referida função para o município analisado. De toda forma, importa analisar com cautela os gastos de referido município, uma vez que é o segundo município do estado a apresentar maior gasto por habitante na função Gestão Ambiental.

5 CONSIDERAÇÕES FINAIS

Em virtude da crise sanitária contemporânea ao período analisado, é plausível prever que despesas adicionais fossem necessárias para que o bem-estar social fosse mantido. Assim, esta pesquisa debruçou-se na análise de referidos gastos e, a partir de então, propôs um método em 4 estágios para aferição de *red flags*.

Pela aplicação do método até o estágio 3, constatou-se que 10 das 26 funções¹³ analisadas (as funções Legislação e Essencial à Justiça não foram analisadas) apresentaram crescimento ou guardavam grande volume de recursos e, por sua relevância, deveriam ser observadas com cautela. A título ilustrativo, foram seguidas ao estágio 4 do método o aprofundamento de 3 das 10 funções eleitas. Referido aprofundamento, permitiu identificar, por função, os municípios que demandam mais atenção, bem como quais as subfunções daqueles municípios devem ser examinadas com maior profundidade.

Assim, entende-se que a pesquisa alcançou seu objetivo ao propor e aplicar um método de detecção de *red flags* para auxiliar na identificação de indícios de fraude nos gastos públicos municipais. Esta pesquisa contribui com os campos acadêmico e profissional, pois ao primeiro acrescenta um conhecimento novo, representado uma combinação de métodos para solução de um problema; com relação ao segundo, disponibiliza aos agentes públicos, em especial aos analistas e auditores de Tribunais de Contas, uma metodologia que permite a detecção de *red flags* sem aplicação de técnicas de *machine learning*.

¹³ Funções previstas na Portaria n°. 42/1999.

REFERÊNCIAS

Andvig, J. C. (2008). Corruption and armed conflicts: some stirring around in the governance soup (No. 2008-3). *Economics Discussion Papers*.

Auriol, E., Straub, S., & Flochel, T. (2016). Public procurement and rent-seeking: the case of Paraguay. *World Development*, 77, 395-407.

Cuiabano, S. M., Leandro, T., Oliveira, G. A. S., & Bogossian, P. (2014). Filtrando cartéis: a contribuição da literatura econômica na identificação de comportamentos colusivos. *Revista de Defesa da Concorrência*, 2(2), 43-63.

Decreto legislativo nº543, de 3 de abril de 2020. Reconhece, para os fins do disposto no art. 65 da lei complementar federal nº101, de 4 de maio de 2000, a ocorrência de estado de calamidade pública, nos termos da solicitação do governador do estado, encaminhada por intermédio da mensagem nº8.502, de 1.º de abril de 2020. Fortaleza, CE. Recuperado em 24 setembro, 2020: <https://www.cge.ce.gov.br/wp-content/uploads/sites/20/2020/04/Decreto-Legislativo-n%C2%BA-543-03-04-2020-Calamidade-P%C3%BAblica.pdf>

Decreto legislativo nº544, de 3 de abril de 2020. Reconhece, para os fins do disposto no art. 65 da lei complementar nº101, de 4 de maio de 2000, a ocorrência do estado de calamidade pública no município de fortaleza, nos termos da solicitação do sr. prefeito, encaminhada por intermédio da mensagem nº 001, de 30 de março de 2020. Fortaleza, CE. Recuperado em 24 setembro, 2020: <https://www.taxpratico.com.br/pagina/decreto-legislativo-n-544-de-3-de-abril-de-2020>

Decreto nº 33.555, de 28 de abril de 2020. Ratifica, para os fins que estabelece, a declaração da ocorrência de calamidade pública em todo o estado do ceará, em razão da pandemia da Covid-19, doença infecciosa viral – cobrade: 1.5.1.1.0, e dá outras providências. Fortaleza, CE. Recuperado em 24 setembro, 2020: <https://www.taxpratico.com.br/pagina/decreto-n-33555-de-28-de-abril-de-2020>

Decreto nº33.510, de 16 de março de 2020. Decreta situação de emergência em saúde e dispõe sobre medidas para enfrentamento e contenção da infecção humana pelo novo coronavírus. Fortaleza, CE. Recuperado em 24 setembro, 2020: <https://www.ceara.gov.br/wp-content/uploads/2020/04/DECRETO-N%C2%BA33.510-de-16-de-mar%C3%A7o-de-2020.pdf>

Dycus, Dennis F. (2001). *Auditing for Fraud*. EUA: ACFE.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.

Gupta, S., Davoodi, H., & Alonso-Terme, R. (2002). Does corruption affect income inequality and poverty?. *Economics of governance*, 3(1), 23-45.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Hawkins, D. M. (1980). *Identification of outliers (Vol. 11)*. London: Chapman and Hall.

Lei Complementar nº 101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. Brasília, DF. Recuperado em 24 setembro, 2020: http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm

Lei nº 13.979, de 6 de fevereiro de 2020. Dispõe sobre as medidas para enfrentamento da emergência de saúde pública de importância internacional decorrente do coronavírus responsável pelo surto de 2019. Brasília, DF. Recuperado em 24 setembro, 2020: <https://www.in.gov.br/en/web/dou/-/lei-n-13.979-de-6-de-fevereiro-de-2020-242078735>

Lei nº 8.666, de 21 de junho de 1993. Regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da Administração Pública e dá outras providências. Brasília, DF. Recuperado em 24 setembro, 2020: http://www.planalto.gov.br/ccivil_03/leis/l8666cons.htm

Lima, R. M. D. (2017). *Ineficiência técnica das prefeituras do estado da Paraíba na alocação de recursos para aquisição de combustíveis*. Dissertação. Programa de Pós-Graduação em Economia do Setor Público da Universidade Federal da Paraíba (UFPB).

López-Iturriaga, F. J., & Sanz, I. P. (2018). Predicting public corruption with neural networks: An analysis of spanish provinces. *Social Indicators Research*, 140(3), 975-998.

Magro, C. B. D., & Cunha, P. R. D. (2017). Red flags na detecção de fraudes em cooperativas de crédito: percepção dos auditores internos. *Revista Brasileira de Gestão de Negócios*, 19(65), 469-491.

Mauro, P. (1995). Corruption and growth. *The quarterly journal of economics*, 110(3), 681-712.

Niessen, M. E. K., Paciello, J. M., & Fernandez, J. I. P. (2020, April). Anomaly Detection in Public Procurements using the Open Contracting Data Standard. In *2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG)* (pp. 127-134). IEEE.

Oliveira, C. D., De Caroli, A. A., de Souza Amaral, A., & Vilca, O. L. (2014). Detecção de fraudes, anomalias e erros em análise de dados contábeis: Um estudo com base em outliers. *Revista Eletrônica do Departamento de Ciências Contábeis & Departamento de Atuária e Métodos Quantitativos (REDECA)*, 1(1), 102-127.

Portaria nº 188, de 3 de fevereiro de 2020. Declara Emergência em Saúde Pública de importância Nacional (ESPIN) em decorrência da Infecção Humana pelo novo Coronavírus (2019-nCoV). Brasília, DF. Recuperado em 24 setembro, 2020: <https://www.in.gov.br/en/web/dou/-/portaria-n-188-de-3-de-fevereiro-de-2020-241408388>

Portaria no 42, de 14 de abril de 1999. Atualiza a discriminação da despesa por funções de que tratam o inciso I do § 1o do art. 2o e § 2o do art. 8o, ambos da Lei no 4.320, de 17 de março de 1964, estabelece os conceitos de função, subfunção, programa, projeto, atividade, operações especiais, e dá outras providências. Brasília, DF. Recuperado em 20 setembro, 2020: http://www.orcamentofederal.gov.br/orcamentos-anuais/orcamento-1999/Portaria_Ministerial_42_de_140499.pdf/

Silva, M. F. G, Freitas, F. G., & Bandeira, A. C. (2001). *How Does Corruption Hurt Growth? Evidences about the Effects of Corruption on Factors Productivity and Per Capita Income*. São Paulo: Fundação Getúlio Vargas.

Souza, F. P. de, & Louzada, F. R. da (2020). Detectando conluio em compras governamentais: Uma abordagem utilizando Red Flags e a Teoria Dempster-Shafer. *Revista da CGU*, 12(21), 80-99.

Tanzi, V., & Davoodi, H. (1998). Corruption, public investment, and growth. In: Shibata, H., & Ihuri, T. (Ed.). *The Welfare State, Public Investment, and Growth*. Tokyo: Springer Japan, 41-60.

Tas, B. K. O. (2017). *Collusion Detection in Public Procurement with Limited Information*. Available at SSRN 2929222.

Tóth, B., Fazekas, M., Czibik, Á., & Tóth, I. J. (2014). *Toolkit for detecting collusive bidding in public procurement*. With examples from Hungary.

Wang, Y. (2016). *Detecting Fraud in Public Procurement* (Doctoral dissertation, The Graduate School, Stony Brook University: Stony Brook, NY.).

APÊNDICE 1

Comandos em Python para regressão linear simples

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

plt.rc('patch',linewidth=2)
plt.rc('axes', linewidth=2, labelpad=10)
plt.rc('xtick.minor', size=4, width=2)
plt.rc('xtick.major', size=8, width=2, pad=8)
plt.rc('ytick.minor', size=4, width=2)
plt.rc('ytick.major', size=8, width=2, pad=8)
# plt.rc('text', usetex=True)
plt.rc('font', serif='Times New Roman', size=4)

df_dados_gastos = pd.read_csv('/home/xx/Área de Trabalho/dados_tce/Dados_V7.csv')
df_dados_pop = pd.read_csv('/home/xx/Área de Trabalho/dados_tce/pop.csv')
df_pop_gastos = pd.merge(df_dados_gastos, df_dados_pop, left_on='município',
right_on='município')
df_pop_gastos = df_pop_gastos.filter(['cod','ano','mês','função','subfunção','despesa',
'valor_liquidado','município','pop','partido'])
df_pop_gastos['valor_liquidado'] =
df_pop_gastos['valor_liquidado'].str.replace(',','.').astype(float)
df_pop_gastos.to_csv('dados_integrados.csv',index=False)
df_scatter = df_pop_gastos
df_scatter = df_scatter.sort_values(by='pop', ascending=True)

#####FILTROS#####
XX
#####FILTROS#####

x = np.log10(df_scatter['pop'].to_numpy())
y = np.log10(df_scatter['valor_liquidado'].to_numpy())

slope, intercept, r_value, p_value, std_err = stats.linregress(x,y)
print('slope={}'.format(slope))
```

```
print('r={}'.format(r_value))
print('intercept={}'.format(intercept))
print('p_value={}'.format(p_value))

# plt.xscale('log')
# plt.yscale('log')

plt.xlabel('pop', fontsize=14)
plt.ylabel('gasto', fontsize=14)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.plot(x, y, 'ok', alpha=0.5)
plt.plot(x, intercept + slope*x, 'r', label='fitted line')
municipios = df_scatter['municipio'].to_numpy()
for i, txt in enumerate(municipios): plt.annotate(txt, (x[i], y[i]))
plt.show()
```


A QUALIDADE E EXPANSÃO DO ENSINO SUPERIOR CONTRIBUI PARA O *OVEREDUCATION*? EVIDÊNCIAS PARA O CEARÁ

Celina Oliveira*

Francisca Zilania Mariano**

Jair Andrade Araújo****

RESUMO

Quais condicionantes induzem o indivíduo a ser *overeducated* e a receber salário abaixo da média? Com dados do Censo de Educação Superior e da RAIS e aplicando o método *Logit* destaca-se que a graduação nas áreas de Agricultura e Veterinária, Ciências, Matemática e Computação, Humanidades e Artes, Serviços, Ciências Sociais, Negócios e Direito ofertados por instituições públicas e as três últimas áreas ofertadas por instituições privadas, bem como, a expansão de vagas no ensino superior fora da capital Fortaleza induzem a uma maior probabilidade do formado ser *overeducated*. Quanto ao desvio salarial, a expansão do ensino na capital contribui positivamente.

Palavras-chave: *Overeducation*; Desvio Salarial; Expansão da Educação Superior.

ABSTRACT

What determinants induce the individual to be overeducated and to receive below of average earnings? With data from the Censo de Educação Superior and RAIS and applying the Logit method it is highlighted that the graduation in the areas of Agriculture and Veterinary; Science, Mathematics and Data Processing; Humanities and Arts; Services; Social Sciences; Business and Law offered by public institutions and the last three areas offered by private institutions, as well as, the expansion of higher education vacancies outside the capital Fortaleza induce a greater probability of the graduates being overeducated. As regards wage deviation, the expansion of education in the capital contributes positively.

Keywords: Overeducation; Wage Deviation; Expansion of Higher Education.

Classificação JEL: I23, J24, J31

Área 2 – Economia Social

* Bolsista Pesquisadora pela FUNCAP no Centro de Análise de Dados e Avaliação de Políticas Públicas (CAPP) do Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE). E-mail: oli.celina@gmail.com. Tel:(85) 997707910.

** Professora da Universidade Federal do Ceará (UFC/Campus de Sobral). E-mail: zilania@ufc.br. Tel: (85) 988177364.

**** Professor do Programa de Pós-Graduação em Economia Rural da Universidade Federal do Ceará (PPGER/UFC). E-mail: jairandrade@ufc.br. Tel: (85) 996807584.

1 INTRODUÇÃO

Nas últimas décadas o governo federal brasileiro tem aplicado planos e programas educacionais de nível superior¹ com o objetivo de estimular o desenvolvimento socioeconômico em cidades que apresentavam os piores índices de desigualdade, por meio da criação de novos campi universitários em municípios que não eram atendidos pelo ensino superior, criação de novos cursos e, conseqüentemente, novas ofertas de vagas. Esse movimento, ficou conhecido como interiorização das universidades.

Os referidos planos e programas acabaram por gerar uma expansão do ensino superior que, por sua vez, trouxe reflexos sobre o mercado de trabalho. Como aponta Rocha et al. (2017), os últimos anos o Brasil tem apresentado um aumento na taxa de ocupação de pessoas com nível superior. Dada essa questão e o que foi colocado por Pritchett (2001) ao utilizar uma amostra de diversos países, encontrou que não há associação entre o aumento do capital humano atribuível ao aumento da escolaridade da força de trabalho e a taxa de crescimento da produção por trabalhador. O autor coloca três possíveis explicações para esse resultado. Primeiro, a escolaridade criou habilidades cognitivas e essas habilidades têm sido usufruídas em atividades não compatíveis, de modo que o capital humano adquirido foi para atividades que propiciam lucros privados, mas socialmente improdutivas. Segundo, há um descompasso entre a taxa de crescimento da demanda por mão-de-obra qualificada e a oferta da mesma. Em terceiro, em alguns países, o sistema educacional não foi eficaz, não gerando habilidades e, conseqüentemente, não gerando produtividade.

A partir dos resultados de Pritchett (2001) e do aumento na taxa de ocupação de pessoas com nível superior, procura-se, na presente pesquisa, averiguar se a expansão e a qualidade do ensino superior têm contribuído para o descasamento entre a formação de nível superior e a ocupação no mercado de trabalho. A não adequação da ocupação com o nível de escolaridade é definida como *overeducation*, ou seja, quando a escolaridade de um indivíduo for maior do que a exigida pela sua ocupação atual.

Sendo assim, o presente trabalho visa contribuir com literatura sobre os determinantes do *overeducation* no mercado de trabalho considerando todas as IES, públicas e privadas, identificando fatores acadêmicos e de expansão do ensino superior. Para tanto, faz-se uso de uma base de dados inédita, formada pelo cruzamento do Censo de Educação Superior de 2012 identificado e a Relação Anual de Informações Sociais (RAIS) de 2015, o qual permitiu observar o *status* dos formados em 2012 no mercado de trabalho três anos após a sua formação. Entre os demais determinantes, inclui-se fatores individuais como idade, sexo e cor; informações sobre educação, como por exemplo, se o formado é oriundo de escola pública, a qualidade do curso de nível superior, as áreas de formação e informações sobre o emprego atualmente ocupado pelo formado. Além disso, objetiva-se analisar o efeito desses determinantes sobre o desvio dos salários comparados à média salarial de cada curso, ou seja, pretende-se investigar os fatores que levam os formados a receberem um salário abaixo da média salarial do seu curso de formação.

Cabe destacar que a análise conduzida apenas para o Estado do Ceará. Segundo o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), autarquia vinculada ao Ministério da Educação, mostram que entre 2003 e 2015 o número de oferta de

¹ Plano Nacional de Educação (PNE): Lei Nº 10.172 de 9 de janeiro de 2001 e Lei Nº 13.005 de 25 de Junho de 2014. A transformação dos Centros Federais de Educação Tecnológica em Institutos Federais de Educação de Ciência e Tecnologia (IFETs): Lei Nº 11.892 de 29 de dezembro de 2008. Fundo de Financiamento Estudantil (FIES): Lei Nº 10.260 de 12 de julho de 2001. Universidade para Todos (PROUNI): Lei Nº 11.096 de 13 de janeiro de 2005. E o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades (REUNI): Decreto Nº 6.096 de 24 de abril 2007.

vagas no ensino superior aumentou 87,5% no Brasil. Entre os estados da região Nordeste, o Ceará apresentou em 2015 o maior aumento na distribuição da oferta, 3,28 p.p. em relação a 2003. Além disso, em parceria com a Secretaria de Educação do Ceará, foi possível a aquisição do Censo de Educação Superior identificado apenas para este Estado.

Além dessa introdução, o artigo está dividido em mais cinco seções, literatura empírica, procedimentos metodológicos adotados, bem como, uma descrição da base de dados utilizada. Após isso, há uma discussão da incidência do *overeducation*, em seguida os resultados empíricos e, por fim, as considerações finais.

2 LITERATURA EMPÍRICA

Overeducation é definido como um descompasso entre o nível de educação alcançado e o nível de educação exigido pelo emprego disponível (FREEMAN, 1976; TSANG; LEVIN, 1985). Estudos sugerem que a situação de *overeducation* existe em muitos países, tanto desenvolvidos, quanto em desenvolvimento.

Frenette (2004) ao realizar um estudo para o Canadá descobriu que 30% dos diplomados eram *overeducation*. Para países em desenvolvimento ou recém-industrializados, Quinn e Rubb (2006) descobriram que no México a incidência de *overeducation* é de 17,2%. Reis (2017) e Marioni (2018) encontraram para o mercado de trabalho brasileiro uma proporção de indivíduos *overeducated* em torno de 25%.

Freeman (1976) elaborou um dos primeiros estudos empíricos sobre *overeducation* para os anos 70 nos Estados Unidos, no qual encontrou retornos decrescentes para a educação, levantando questões em relação ao excesso de produção de graduados. Duncan e Hoffman (1981) estimaram os impactos da diferença entre os anos de estudos possuídos por um trabalhador e os requeridos pela sua ocupação sobre os rendimentos do trabalhador. Como resultado encontraram o excesso de escolaridade possui um efeito positivo sobre os salários, no entanto, esse efeito é duas vezes menor caso a ocupação exigisse o nível de escolaridade possuído pelo trabalhador.

Seguindo os trabalhos pioneiros, grande parte da literatura tem a preocupação de averiguar os efeitos desse fenômeno sobre os salários. Hartog (2000) analisou para Holanda, Reino Unido, Estados Unidos, Espanha e Portugal. Mcguinness (2006), Leuven e Oosterbeek (2011) utilizaram uma versão estendida da equação salarial Minceriana. Nieto e Ramos (2017) envolvendo a Espanha e Diaz e Machado (2008), Reis (2017) e Marioni (2018) para o Brasil. Todos estudos apontaram que em geral os retornos do *overeducation* são positivos sobre os salários, mas menores que os retornos quando se considera o *match* entre educação e ocupação.

Em relação aos efeitos negativos gerados pela existência de *overeducation*, Tsang e Levin (1985) discute como essa situação pode afetar de forma negativa a produtividade individual. Segundo os autores, trabalhadores com mais educação do que o seu trabalho exige, podem apresentar um comportamento contraproducente no local de trabalho por estarem insatisfeitos com seus empregos. Groot (1996) compartilha da mesma ideia. Para os trabalhadores *overeducated* é possível que estes obtenham rendimentos inferiores aos obtidos por trabalhadores que desempenham atividades condizentes com seu nível de formação, desmotivando-o a melhorar sua capacitação, tendo em vista que o investimento em qualificação não possui o retorno esperado na forma de salário. Portanto, o declínio da produtividade dos trabalhadores devido ao *overeducation* tem sido associado a uma menor produtividade e a baixa satisfação do trabalho (GROOT; DEN BRINK, 2000).

Verhaest e Omeij (2006) constataram que trabalhadores *overeducated* são menos satisfeitos com seus empregos, tendem a mudar de emprego com mais frequência e são menos

motivados a participar de treinamentos. Pode-se dizer então que a situação de *overeducation* tende a ter um impacto negativo na atitude dos indivíduos em relação ao seu trabalho, podendo resultar em menor produtividade.

Todas essas consequências passam a afetar a lucratividade das empresas gerando assim um problema a ser enfrentado pelo setor privado. O setor público também absorve seus efeitos negativos por, também, ser empregador e, principalmente, por ser o principal vetor de financiamento da educação, o que gera uma preocupação do retorno do investimento governamental em educação.

Com relação aos determinantes do *overeducation*, Frenette (2004) expõe determinantes socioeconômicos para o Canadá. Entre eles, o fato de a graduação ser acompanhada por um programa de estágio, possuir mestrado, ser formado em área que envolve prestação de serviços ao consumidor, ter experiência de trabalho antes da graduação, condições econômicas locais como alta taxa de desemprego parecem deixar os indivíduos mais propensos a ser superqualificados. Fatores como: sexo, ser casado e ter filhos não apresentaram evidências significantes que influencie indivíduos a aceitar trabalhos de nível mais baixo que sua escolaridade permite.

Na mesma linha, Morano (2014) analisa os determinantes do *overeducation* para o mercado de trabalho italiano considerando trabalhadores com mestrado. Os resultados apontam uma correlação negativa da idade e o tamanho da empresa com a situação de *overeducation*. Indivíduos casados, ser do sexo masculino, ser estrangeiro, estar em um emprego em tempo parcial e temporário e do setor de serviços são fatores que influenciam de forma positiva a propensão de ser *overeducated*. Além disso, indivíduos formados em áreas de um grau mais científico (por exemplo, engenharia, ciências científicas) ou um grau muito específico (por exemplo, medicina, arquitetura) são menos propensos a ser *overeducated*. O autor documenta ainda que as condições do mercado de trabalho afetam o a situação de *overeducation* como exemplo, o efeito positivo do desemprego.

Estudos para o Brasil revelam que o término do ensino superior (efeito-diploma), experiência no mercado de trabalho e residir no meio rural aumentam as chances de os ocupados estarem na situação de *overeducation*. No entanto, o indivíduo enquanto do sexo feminino, branco, residente em regiões metropolitanas e estar sindicalizado influencia de forma negativa as chances de os mesmos estarem nessa situação (CAVALCANTI; CAMPOS; SILVEIRA NETO, 2010).

Ainda para o Brasil, Anegues *et al.* (2018) considerando uma base de dados de egressos da Universidade Federal da Paraíba (UFPB) cruzada com dados sobre mercado de trabalho da Relação Anual de Informações Sociais (RAIS) investigou quais fatores afetam a probabilidade de um indivíduo estar em uma ocupação com nível de instrução exigido abaixo da sua formação superior. Como resultado, destacou o efeito da área de formação dos egressos. As áreas de humanidades, artes e sociais aplicadas apresentam maior probabilidade de os indivíduos estarem sob *overeducation*, bem como uma menor chance de deixar essa condição durante alguns anos após a conclusão do curso.

Nesse mesmo caminho, literaturas tem destacado que o fenômeno de *overeducation* pode ser mais comum em determinadas áreas de formação. As áreas ligadas a ciências e tecnologia podem fornecer maiores níveis de “habilidades produtivas” do que as áreas de humanidades, pois o currículo desses cursos enfatiza mais a aquisição de habilidades específicas requeridas pelas ocupações no mercado de trabalho (REIMER; NOELKE; KUCEL, 2008). Tal hipótese vem sendo corroborada pelas evidências empíricas. Dolton e Vignoles (2000) e Dolton e Silles (2008) analisaram os determinantes do *overeducation* no mercado de trabalho entre indivíduos graduados no Reino Unido. Ambos os trabalhos encontram que

trabalhadores formados nas áreas de humanidades e artes apresentaram maior probabilidade de estarem sobreeducados em comparação com outros campos de estudo.

Diante disso, o objetivo do presente estudo consiste em fornecer evidências empíricas sobre os determinantes do *overeducation* no Ceará em 2015, destacando o papel da expansão do ensino superior nos últimos anos, bem como, as áreas de formação ofertadas, diferenciando entre público e privada, a qualidade do curso, dentre outras características individuais e da firma na qual o indivíduo possui vínculo.

3 PROCEDIMENTOS METODOLÓGICOS

3.1 Dados

O banco de dados utilizado foi gerado a partir da combinação de informações do Censo da Educação Superior e da Relação de Informações Sociais (RAIS). Os dados sobre ensino superior são divulgados anualmente pelo INEP fornece informações sobre as instituições de educação superior (IES), seus cursos de graduação presencial ou a distância, cursos sequenciais, vagas oferecidas, inscrições, matrículas, vagas, dados de financiamento estudantil, ingressantes e concluintes e informações sobre docentes nas diferentes formas de organização acadêmica e categoria administrativa que ofertam cursos de graduação e sequencias de formação específica, além de informações de seus alunos e docentes. O Censo ainda contribui para o cálculo de indicadores de qualidade como o Índice Geral de Cursos (IGC), que classifica as instituições em uma escala de 1 a 5. Este indicador será utilizado na presente pesquisa como uma variável de qualidade da Instituição de Ensino.

A RAIS traz informações referente ao mercado de trabalho ocupado pelos indivíduos que constam no Censo da Educação Superior. Dentre as informações, constam as características individuais dos trabalhadores (raça, sexo, grau de instrução, idade, tipo de admissão, salário anual) e informações sobre as ocupações e sobre as empresas, como seu porte, classificação de atividade e natureza jurídica.

A amostra é composta por indivíduos que concluíram sua graduação nas IES do Ceará no ano de 2012 (19.272 pessoas) e estavam presente na RAIS em 2015, totalizando 8.981 observações, o equivalente a 46,6%. O objetivo do cruzamento é observar se os formados em 2012 estavam trabalhando três anos após sua formação e, se o cargo ocupado correspondia com um cargo de nível superior. Essa amostra foi construída considerando apenas os formados empregados com vínculo ativo, que ocupam cargos segunda a Classificação Brasileira de Ocupações de 2002 (CBO 2002) com exceção dos militares, dirigentes do poder público e caciques. O indicador de *overeducation* foi construído a partir da abordagem do *Job Analysis* que consiste em utilizar uma classificação ocupacional que seguem os graus de escolaridade necessários para as ocupações, na qual é definida por analistas de trabalho. Este método foi utilizado por Hartog (2000), McGuinness (2006), Morano (2014) e Annegues et al. (2018). Segundo Hartog (2000), é o método mais atraente porque são baseadas na tecnologia do trabalho associada a cada ocupação e está menos sujeita a vieses individuais (DOLTON; SILLES, 2008; LEUVEN; OOSTERBEEK, 2011). Seguindo essa abordagem, uma das variáveis presente na RAIS é a ocupação organizada no sistema da Classificação Brasileira de Ocupações (CBO) de 2002. Segundo o Ministério do Trabalho e Emprego (MTE) a ocupação é a agregação de empregos ou situações de trabalho similares quanto às atividades realizadas.

Sendo assim, CBO 2002 agrega os empregos por habilidades cognitivas comuns exigidas no exercício de um campo de trabalho².

O nível mais agregado da CBO 2002 são os grandes grupos formado por dez conjuntos, agregados por nível de competência e similaridade nas atividades executadas. A CBO 2002 toma como base, com algumas adaptações, a metodologia da classificação e agregação internacional de informações ocupacionais de 1988 a International Standard Classification of Occupations (ISCO 88) que utiliza a escolaridade como critério para determinar o nível de competência. Os graduados cuja ocupação na CBO tenha nível de escolaridade exigido menor que o ensino superior são considerados sobreeducados e, assim, a dummy de overeducation será igual a um.

3.2 Estratégia Empírica

Quando se pretende identificar os determinantes do *overeducation* e do desvio salarial deve-se levar em consideração o de viés de seletividade amostral. Ao fazer o cruzamento do Censo de Educação Superior de 2012 com a Relação Anual de Informações Sociais de 2015, as informações coletadas são fornecidas apenas para aqueles que estavam no mercado de trabalho formal no período da pesquisa, ou seja, os salários e as ocupações observados estão relacionados com a decisão de um indivíduo trabalhar ou não e esta pode estar correlacionada a fatores não observáveis. Com isso, a consideração somente de pessoas que estejam trabalhando pode produzir estimadores viesados (WOOLDRIDGE, 2008).

Para solucionar o problema do possível viés de seletividade amostral, será aplicado o procedimento de Heckman (1979) em dois estágios, o qual é formado por duas equações; uma que descreve a participação no mercado de trabalho em 2015 inserindo todos os formados em 2012, os que foram identificados no mercado de trabalho e os que não foram, e outra do tipo Logit para encontrar os determinantes do *overeducation* e do desvio salarial no mercado de trabalho formal. O primeiro estágio consiste em estimar a seguinte equação de participação:

$$L = \beta_i X_i + \mu_i \quad (1)$$

onde L é a variável que designa a participação no mercado de trabalho e X_i representa um conjunto de variáveis que explicam a participação no mercado de trabalho. Como L não é observado, utiliza-se uma variável *dummy* definida como segue:

$$\begin{aligned} L_i &= 1 \text{ se } L_i^* > 0 \\ L_i &= 0 \text{ se } L_i^* \leq 0 \end{aligned} \quad (2)$$

Neste caso, L assume o valor 1 para representar o indivíduo que está trabalhando e 0, caso contrário. Após a estimação do modelo binário com a hipótese probit, o segundo procedimento para aplicar o método de Heckman (1979) consiste em estimar a equação dos determinantes de *overeducation* e desvio salarial, levando em consideração o viés de seleção amostral.

Será estimado dois modelos, no primeiro modelo, estima-se os determinantes do *overeducation*, em que Y_i é igual a 1 se o indivíduo ocupa um cargo que exige uma escolaridade abaixo do nível superior e zero, caso contrário. No segundo modelo, estima-se os determinantes

² O MTE define o sistema de classificação da CBO 2002 como um documento normalizador do reconhecimento para fins classificatórios, sem função de regulamentação profissional, da nomeação e da codificação dos títulos e conteúdos das ocupações do mercado de trabalho brasileiro para fins estatísticos de registros administrativos, censos populacionais e outras pesquisas domiciliares.

do desvio salarial, em que Y_i será igual a 1 se o formado receber um salário abaixo da média do seu curso de graduação e zero, caso contrário. A equação 2 traz a especificação dos modelos.

$$Y_i = \beta_0 + \beta_1 Individual_i + \beta_2 Educação_i + \beta_3 Expansão_i + \beta_4 Emprego + \beta_5 \varphi_i + \varepsilon_i \quad (3)$$

Em ambos os modelos utiliza-se as mesmas variáveis explicativas, compostas pelos grupos de variáveis de características individuais como: idade, cor e sexo; educacionais como: ensino médio de escola pública, qualidade da graduação, áreas dos cursos segundo a classificação do INEP divididas entre instituições privadas e públicas; expansão do ensino superior representados pelo logaritmo da diferença entre o número de vagas ofertadas no ensino superior por município entre os anos de 2003 e 2015 divididas em oferta para a capital Fortaleza, Região Metropolitana sem o município de Fortaleza e para os municípios do interior. Por fim, considerou-se ainda as características do emprego atual do graduado, como por exemplo, se é o primeiro emprego, se é emprego temporário, a experiência no emprego atual e a localização

geográfica dentro do Estado. A função $\varphi_i = \frac{\phi(\beta_i X_i)}{\Phi(\beta_i X_i)}$ é denominada razão inversa de Mills (IMR) e representa a variável que segundo Heckman (1979) é utilizada como explicativa na regressão principal para eliminar o viés de seletividade amostral, onde a significância desta função indicará a presença de viés de seletividade amostral. Todas as variáveis estão descritas no Quadro 1.

Quadro 1: Descrição das Variáveis dos Modelos

Modelo 1 - Variável Dependente: Overeducation	
Y_i	= 1 se está na situação de <i>Overeducation</i> e zero, c.c..
<i>Overeducation</i>	
Modelo 2 - Variável Dependente: Desvio Salarial	
Y_i	= Desvio Salarial 1 se o salário for abaixo da média do curso de formação e zero, c.c..
Variáveis Independentes	
Características Individuais	
Idade	Idade do formado em anos
Branco	1 para raça branca e zero, c.c.
Homem	1 para se for do sexo masculino e zero, c.c.
Características Educacionais	
Bolsa	1 recebe alguma bolsa da IES seja estágio, extensão, monitoria ou pesquisa e zero, c.c.
IGC	Indicador de qualidade da Instituição de Ensino Superior.
Agr. Veter. Púb.	1 se o formado possui graduação na área de Agricultura e Veterinárias em uma IES Pública e zero, c.c.
Serv. Púb.	1 se o formado possui graduação na área de Serviços em uma IES Pública e zero, c.c.
Saúde e Bem Estar Social Púb.	1 se o formado possui graduação na área de Saúde e Bem Estar Social em uma IES Pública e zero, c.c.
Ciên. Soc., Neg. Dir. Púb.	1 se o formado possui graduação na área de Ciências Sociais, Negócios e Direito em uma IES Pública e zero, c.c.
Ciên., Mat. Comp. Púb.	1 se o formado possui graduação na área de Ciências, Matemática e Computação em uma IES Pública e zero, c.c.
Educ. Púb.	1 se o formado possui graduação na área de Educação em uma IES Pública e zero, c.c.

Eng., Prod. Const Púb	1 se o formado possui graduação na área de Engenharia, Produção e Construção em uma IES Pública e zero, c.c.
Human. Artes Púb.	1 se o formado possui graduação na área de Humanidades e Artes em uma IES Pública e zero, c.c.
Agr. Veter. Priv.	1 se o formado possui graduação na área de Agricultura e Veterinárias em uma IES Privada e zero, c.c.
Serv. Priv.	1 se o formado possui graduação na área de Serviços em uma IES Privada e zero, c.c.

continua

Saúde e Bem Estar Social Priv.	1 se o formado possui graduação na área de Saúde e Bem Estar Social em uma IES Privada e zero, c.c.
Ciên. Soc., Neg. Dir. Priv.	1 se o formado possui graduação na área de Ciências Sociais, Negócios e Direito em uma IES Privada e zero, c.c.
Ciên., Mat. Comp. Priv.	1 se o formado possui graduação na área de Ciências, Matemática e Computação em uma IES Privada e zero, c.c.
Educ. Priv.	1 se o formado possui graduação na área de Educação em uma IES Privada e zero, c.c.
Eng., Prod. Const. Priv	1 se o formado possui graduação na área de Engenharia, Produção e Construção em uma IES Privada e zero, c.c.
Human. Artes Priv.	1 se o formado possui graduação na área de Humanidades e Artes em uma IES Privada e zero, c.c. Utilizada como base.

Características Expansão do Ensino Superior

Oferta de Vagas Fortaleza	Logaritmo da diferença entre o número de vagas ofertadas no ensino superior em Fortaleza, entre 2003 e 2015.
Oferta de Vagas RMF	Logaritmo da diferença entre o número de vagas ofertadas no ensino superior na Região Metropolitana de Fortaleza, excluindo Fortaleza, entre 2003 e 2015.
Oferta de Vagas Interior	Logaritmo da diferença entre o número de vagas ofertadas no ensino superior no Interior do Ceará, entre 2003 e 2015.

Características do Emprego Atual

RMF	1 se o emprego atual está na Região Metropolitana de Fortaleza e zero, c.c.
Prim. Emp.	1 se o vínculo empregatício atual for o primeiro emprego e zero, c.c.
Emp. Temp.	1 se o vínculo empregatício atual for emprego temporário e zero, c.c.
Exper. Empr. Atual	Logaritmo do Tempo de vínculo empregatício no emprego atual em anos.
IMR	Razão Inversa de Mills

Fonte: Elaboração dos Autores

Nota: c.c. significa caso contrário.

4 EVIDÊNCIAS INICIAIS

A Tabela 1 traz estatísticas que descrevem a amostra considerada na presente pesquisa, indivíduos com nível superior firmados pelas IES do Ceará em 2012 observados no mercado de trabalho formal em 2015. Considerando os indivíduos apenas na situação de *Overeducation*, observa-se que os indivíduos são relativamente jovens com média de idade de 32 anos e, aproximadamente, dois anos mais velhos do que os indivíduos que estão em *match*, aqueles que possuem nível superior e ocupam uma vaga que exige essa escolaridade.

Para o modelo de desvio salarial, aqueles que apresentam um desvio salarial abaixo da média de salário do seu curso de formação são relativamente jovens, com média de idade de 30 anos e, aproximadamente 1 ano mais novo do que aqueles indivíduos que não apresentaram desvio salarial.

Quanto ao indicador de qualidade da instituição de ensino superior, os indivíduos *overeducated* apresentam um IGC um pouco menor que aqueles indivíduos que não estão nessa situação. Já para o desvio salarial, aqueles, que não apresentaram desvio para abaixo da média de salário possuem um IGC relativamente maior.

Com relação as variáveis que representam a expansão do ensino superior nos últimos anos, observa-se que a oferta é em média maior para os que estão em situação de *match* quando essas vagas são ofertadas na capital. Por outro lado, se essas vagas são direcionadas para os demais municípios da RMF e para os municípios do Interior a média de oferta de vagas torna-se maior para o grupo dos *overeducated*. Esse resultado, já aponta para uma primeira evidência de que a expansão do ensino superior nas regiões fora da capital pode induzir a uma situação de *overeducation*.

A oferta de vagas em Fortaleza e da RMF é em média maior no grupo de indivíduos que recebem salário maior ou igual a média de salários do seu curso de formação. Esse resultado já muda quando a oferta se direciona para o interior.

Com relação a experiência no emprego atual, os formados *overeducated* ficam em média um pouco mais de tempo no emprego atual, em relação aos indivíduos em *match*. No modelo de desvio salarial, os indivíduos que não apresentaram desvio passam mais tempo no emprego atual que aqueles que possuem desvio salarial.

Tabela 1: Evidências Iniciais: estatística descritiva

Variáveis	Overeducation		Match		Desvio Salarial		Não Desvio	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
Idade	31,63	7,09	30,75	6,52	30,7	6,31	31,31	7,04
IGC	2,75	0,45	2,79	0,44	2,75	0,43	2,81	0,46
Oferta de Vagas Fortaleza	6,43	5,78	7,7	5,5	6,88	5,71	7,69	5,5
Oferta de Vagas RMF	0,5	1,77	0,28	1,35	0,29	1,38	0,4	1,59
Oferta de Vagas Interior	1,55	3,13	1,15	2,77	1,49	3,07	1,08	2,7
Exper. Empr. Atual	3,3	1,23	2,94	1,07	2,98	1,12	3,12	1,15
Proporção em (%)								
Branco	Não	72,17	69,12	70,02	70,12			
	Sim	27,83	30,88	29,98	29,88			
Sexo	Feminino	59,48	59,73	63,04	56,59			
	Masculino	40,52	40,27	36,96	43,41			
Ensino Médio	Privado	86,66	91,61	88,39	91,58			
	Público	13,34	8,39	11,61	8,42			
Bolsa	Não	95,96	92,68	93,33	94,03			
	Sim	4,04	7,32	6,67	5,97			
RMF	Não	31,62	26,05	33,22	22,88			
	Sim	68,38	73,95	66,78	77,12			
Primeiro Emprego	Não	90,84	78,76	82,68	82,37			
	Sim	9,16	21,24	17,32	17,63			
Emprego Temporário	Não	99,86	99,79	99,81	99,81			
	Sim	0,14	0,21	0,19	0,19			

Fonte: Elaboração dos Autores

Considerando as variáveis qualitativas, grande parte dos *overeducated* declaram-se não brancos, cerca de 72%. Essa proporção diminui um pouco para o grupo do *match*. A grande participação dos não brancos é verificada também no modelo de desvio salarial.

As mulheres estão em maior participação no grupo dos *overeducated* com 59%, esse mesmo percentual se repete no grupo de *match*. Destaca-se ainda, a grande participação das mulheres no grupo de trabalhadores que possuem desvio salarial abaixo da média de salário do curso de formação.

No grupo dos *overeducated* apenas 13% cursaram seu ensino médio em escola pública, esse percentual diminui para 8% no grupo de *match*. Com relação ao modelo de desvio salarial,

cerca de 12% dos que recebem salário abaixo da média cursaram seu ensino médio em escola pública, esse percentual diminui para quase 8% no grupo dos que recebem salário igual ou acima da média de salário do curso de formação.

Apenas 4% dos *overeducated* receberam algum tipo de bolsa universitária, esse percentual aumenta para quase 7% no grupo de *match*. Já no grupo que recebe salário abaixo da média quase 7% recebem bolsa, esse percentual cai quase 1% para o grupo dos que recebem salários acima ou igual a média.

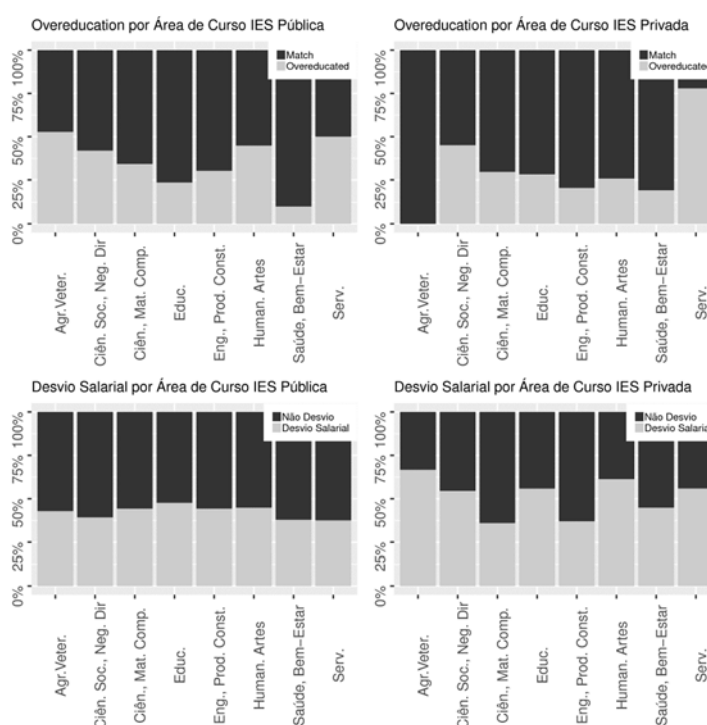
A maioria dos *overeducated* trabalham na RMF, 68,38%. Esse percentual aumenta para 73,95% para o grupo de *match*. No grupo de desvio salarial, grande parte também trabalha na RMF, 66,78%. O mesmo ocorre para o grupo que possui salário acima ou igual, porém em maior proporção, 77,12%.

Apenas 9% dos *overeducated* estão no primeiro emprego. Já no grupo de *match* o percentual de formados no primeiro emprego aumenta para 21,24%. No modelo de desvio salarial, o percentual é cerca de 17% para ambos os grupos. Por fim, para todos os grupos, quase nenhum dos formados estão em vínculos empregatícios temporários.

A Figura 1 contém quatro gráficos com a distribuição dos formados em cada área de curso de ensino superior dividido por IES públicas e privadas para ambos os modelos. A parte cinza das barras representa os indivíduos em situação de *overeducation* ou que possuem desvio salarial abaixo da média, enquanto que a porção mais escura representa aqueles que possuem um *match* adequado entre ocupação e escolaridade ou recebem salários acima ou igual a média do curso de formação.

Com relação as instituições públicas de ensino a área de Agricultura e Veterinária é a que apresenta maior proporção de formados em situação de *overeducation* no mercado de trabalho cearense em 2015. Outras áreas cuja maior parte dos formados estão em empregos que não exigem nível superior são a área de Serviços e Humanidades e Artes. Já área de Saúde e Bem-Estar Social apresenta a menor parcela de formados *overeducated*.

Figure 1: Distribuição de *Overeducation* e Desvio Salarial por Área de Curso



Fonte: Elaboração dos autores a partir da amostra selecionada.

Quando se observa as IES privadas, há uma mudança nessas proporções. A área de Agricultura e veterinária passa a ter todos os formados empregados em ocupações que exigem o nível superior, caindo para a área de Ciências Sociais, Negócios e Direito a maior proporção de *overeducated*. Dentre as demais áreas, Saúde e Bem-Estar Social e Engenharia, Produção e Construção possuem uma menor proporção de *overeducated*. Comparando os tipos de instituições de ensino superior, o setor privado apresenta maior proporção de *overeducated* nas áreas de Serviços; Saúde e Bem-Estar Social; Educação; Ciências Sociais, Negócios e Direito, são exatamente as áreas que concentram maior parte dos cursos ofertados pelas IES privadas, como por exemplo os curso de Direito, Administração e Pedagogia.

No modelo de desvio salarial e considerando as IES públicas, dos que apresentam esse desvio abaixo da média, grande parte cursam a área de Educação. Já as áreas de Saúde e Bem-Estar Social e Serviços mais da metade recebem salários acima ou igual a média salarial do curso de formação. Ao analisar as IES privadas, há uma mudança. A área de Agricultura e Veterinária apresenta a maior proporção de trabalhadores que recebem salário abaixo da média salarial do seu curso de formação e a área de Ciência, Matemática e Computação apresenta a menor proporção de trabalhadores que possuem desvio salarial para baixo. Comparando as IES públicas e privadas, a primeira apresenta sempre uma maior proporção de trabalhadores que não possuem desvio salarial para baixo. Já as privadas apenas os cursos das áreas de Ciência, Matemática e Computação; Engenharia, Produção e Construção e Saúde apresentam uma maior proporção desses indivíduos.

5 RESULTADOS EMPÍRICOS

Nesta seção são apresentados os resultados empíricos dos modelos, determinantes do *overeducation* e do desvio salarial para o Ceará em 2015, enfatizando o papel das áreas de ensino superior e da expansão da oferta de educação superior ocorrida nos últimos anos. A Tabela 2 expressa os coeficientes com os sinais dos determinantes e o *odds ratio* (razão de chance) para o modelo de *overeducation* e para o modelo de desvio salarial.

Como primeira etapa, especificou-se uma variável binária de participação no mercado de trabalho formal e calculou-se a razão inversa de Mills, utilizando como variáveis explicativas cor, idade, sexo, se cursou ensino médio em escola pública e se a IES de nível superior está na capital. Observa-se que essa variável foi significativa nos dois modelos, indicando que a não inclusão desta variável nas equações fornecerá estimativas viesadas. Dessa forma, é possível afirmar que os indivíduos que estão no mercado de trabalho em 2015 foram pré-selecionados se comparados aos formados que não estão trabalhando no emprego formal, ou seja, apresentam características comuns que levaram a tomar essa decisão de trabalhar.

Com relação aos determinantes de características individuais, quanto maior a idade do indivíduo maior a probabilidade de ele ser *overeducated*. Embora relativamente pequena, a chance de indivíduos estarem em situação de *overeducation* é cerca de 1,3% ao se aumentar um ano de idade. Isso possivelmente reflete o fato de indivíduos mais velhos serem mais aversos ao risco. A tendência é que com o passar do tempo, o acúmulo de experiência no cargo que ocupa torna mais custoso a saída do emprego atual para um que se encaixe no nível de escolaridade adquirido. Os autores Cavalcanti, Campos e Silveira Neto (2010) encontraram o mesmo resultado para o Brasil, onde o passar do tempo, o acúmulo de experiência no mercado de trabalho não parece fazer o indivíduo alcançar postos de trabalho mais apropriados à sua qualificação.

Quanto aos determinantes educacionais, o fato de o indivíduo receber qualquer bolsa de estágio, extensão, monitoria ou pesquisa durante sua graduação diminui as chances de ele

ser *overeducated* em cerca de 47%. O mesmo ocorre se ele cursar sua graduação em uma instituição de boa qualidade avaliado pelo IGC. Em termos percentuais, um aumento de uma unidade no IGC induz uma queda de 20% nas chances de o indivíduo estar em um emprego que exija um nível de escolaridade abaixo do que ele possui. Como apontado por Robst (1995) trabalhadores que frequentaram IES de baixa qualidade tendem a ser mais *overeducated*. Segundo o autor, estudantes com menor capacidade e/ou matriculados em IES de baixa qualidade necessitam de mais anos de escolaridade para adquirir as habilidades necessárias para o trabalho. Além disso, se a área de formação corresponder a área de Saúde e Bem-estar Social do setor público há uma redução de cerca 66% de chances do indivíduo ser *overeducated*. Esse resultado reflete a atual necessidade por profissionais da área da Saúde e está compatível com o apontado por Annegues et al. (2018) para o estado da Paraíba, onde os estudantes das áreas de Saúde e Informática possuem um menor risco de estarem em situação de *overeducation*.

Dentre os fatores educacionais que proporcionam uma maior probabilidade do indivíduo ser *overeducated* tem-se o fato dele ter cursado seu Ensino Médio em escola pública aumentando duas vezes mais as chances de ele estar nessa situação em comparação com os alunos do ensino médio de escolas particulares.

Pode-se destacar ainda uma possível diferença entre os efeitos das instituições de ensino superior públicas e privadas, como por exemplo, enquanto a área de Saúde e Bem estar pública reduz a probabilidade de ser *overeducation*, essa área ofertada pelo ensino privado é não significativa. Já para as áreas de Agricultura e Veterinária e Ciências Matemática e computação públicas aumentam a probabilidade de ser *overeducation* as instituições privadas também não apresentaram efeito. Ademais verifica-se diferenças entre as áreas pertencentes ao mesmo tipo de instituição. Considerando o Ensino Superior Público, se o indivíduo cursar uma graduação nas áreas Agricultura e Veterinária, Serviços, Ciências Sociais, Negócios e Direito e Ciências, Matemática e Computação e Humanidades e Artes tendem a ter maiores probabilidades de serem *overeducated*. Dentre esses cursos, Agricultura e Veterinária aumenta quase 5 vezes e Serviços e Humanidades quase 4 vezes mais as chances de o indivíduo estar nesta situação. Já considerando o Ensino Superior Privado, se graduar nas áreas de Serviços e Ciências Sociais, Negócios e Direito também induz a situação de *overeducation*. Em termos de chaces, se graduar na área de Serviços aumenta onze vezes e cursar Ciências Sociais, Negócios e Direito aumenta quase três vezes mais. Este último resultado tem ido de acordo com o encontrado por Morano (2014) para os graduados italianos e por Annegues et al. (2018) para o estado da Paraíba. Segundo estes autores, graduação em áreas específicas ou que possuem pouca mão-de-obra disponível como as áreas de Ciências, Medicina e Engenharia reduzem a probabilidade de ser *overeducation*, enquanto que formados em Economia, Direito, Ciências Sociais, Ciências Humanas, Artes, Comunicação e Linguagens essa probabilidade aumenta. Para essas últimas áreas, os autores colocam que isto é possível, devido à restrições do mercado de trabalho característico desses campos de atuação e acabem por gerar um excesso de oferta de trabalhadores, de modo que para muitos a saída para conseguir um emprego é ocupar cargos com funções não condizentes com sua instrução.

Com relação aos fatores que representam a expansão do ensino superior, um aumento no número de vagas implicará em um aumento na probabilidade de ser *overeducation*, quando essas vagas se direcionam para as regiões fora da capital Fortaleza. No entanto, uma expansão das vagas sobre a capital tenderá a diminuir o *overeducation*. Em termos de chances percentuais, o aumento no número de vagas na capital reduzirá em 5,6% as chances de o indivíduo estar nesta situação. No entanto, se essa expansão das vagas se localizar nos demais municípios da região metropolitana, há um aumento das chances de ser *overeducation* em 4%. E se a expansão se direcionar para o interior o aumento nas chances de estar nesta situação de 2,4%. Este resultado, se deve possivelmente ao fato, em particular, da Cidade de Fortaleza: a saúde econômica e a composição industrial e de serviços da região, podem determinar em

grande parte a demanda por mão-de-obra, principalmente em áreas específicas Frenette (2004). Esse mesmo resultado foi encontrado por Cavalcanti, Campos e Silveira Neto (2010), onde verificou que as regiões metropolitanas e urbanas, por apresentarem mercados mais amplos e dinâmicos, induzem melhores *matches* entre a formação e a ocupação do indivíduo, reduzindo assim, o risco de ele ser *overeducation*.

As características relacionadas ao emprego, indica que se o indivíduo estiver em seu primeiro emprego, há uma redução nas chances de 40% de ser *overeducated*. Este resultado, pode estar relacionado ao fato de que recém formados são mais propensos a serem capazes de aplicar as habilidades e conhecimentos que eles acabaram de adquirir Morano (2014). Em termos de experiência no emprego atual, à medida que há o aumento desta variável, implicará em um aumento 19,65% de estar na mesma situação de *overeducation*.

Os determinantes considerados envolvem informações de características individuais, educacionais e do emprego atual. Dentre os determinantes para a probabilidade do indivíduo ser *overeducated*, não obteve-se significância estatística na variável Branco, ser Homem, se ocupa um emprego temporário, ou se o curso de nível superior da sua formação for da área de Agricultura e Veterinária; Saúde e Bem-Estar Social; Educação; Ciências, Matemática e Computação; Engenharia, Produção e Construção ofertados pelo setor privado e Educação e Engenharia, Produção e Construção ofertado pelo setor público.

Tabela 2: Determinantes do Overeducation e do Desvio Salarial

	Overeducation		Desvio Salarial	
	Coefficientes	Odds Ratio	Coefficientes	Odds Ratio
	0,0131***	1,0132***	-0,0234***	0,9769***
	(0,0050)	(1,0034;1,0230)	(0,0044)	(0,9682;0,9856)
	0,0198	1,0200***	0,0068	1,0068***
	(0,0567)	(0,9088;1,1312)	(0,0502)	(0,9085;1,1051)
	0,0021	1,0021***	-0,3094***	0,7339***
	(0,0547)	(0,8948;1,1093)	(0,0487)	(0,6384;0,8294)
Púb.	0,7810***	2,1836***	-0,2253*	0,7983***
	(0,1315)	(1,9259;2,4412)	(0,1164)	(0,5701;1,0265)
	-0,6228***	0,5364***	-0,0087	0,9914***
	(0,1170)	(0,3071;0,7658)	(0,0934)	(0,8082;1,1745)
	-0,2329***	0,7922***	0,0489	1,0501***
	(0,0665)	(0,6618;0,9226)	(0,0604)	(0,9317;1,1685)
Púb.	1,5636***	4,7761***	-0,7099	0,4917
	(0,4828)	(3,8299;5,7224)	(0,4366)	(-0,3640;1,3474)
	1,2234*	3,3987***	-0,8876	0,4116
	(0,6722)	(2,0811;4,7163)	(0,6464)	(-0,8553;1,6786)
Estar Púb.	-1,0950**	0,3345	-1,0465***	0,3511
	(0,4579)	(-0,5630;1,2320)	(0,3921)	(-0,4174;1,1197)
	0,9217**	2,5136***	-0,8892**	0,4110
Neg. Dir Púb.	(0,4359)	(1,6592;3,3680)	(0,3858)	(-0,3451;1,1671)
	0,8016*	2,2292***	-0,6879*	0,5026
Comp. Púb,	(0,4540)	(1,3394;3,1190)	(0,4028)	(-0,2869;1,2921)
	0,0980	1,1029**	-0,6578*	0,5180
	(0,4328)	(0,2547;1,9512)	(0,3813)	(-0,2294;1,2654)
	0,4323	1,5408***	-0,6139	0,5412
Const. Púb.	(0,4521)	(0,6547;2,4268)	(0,4000)	(-0,2428;1,3253)
	1,3025**	3,6785***	-0,6217	0,5370

Artes Púb.	Human.	(0,5751)	(2,5514;4,8056)	(0,5363)	(-0,5141;1,5882)
Priv.	Agr. Veter.	-113869	0,00001	0,2745	13159
		-1843448	(-361,3092;361,3092)	-13150	(-1,2616;3,8933)
	Serv. Priv.	2,4030***	11,0567***	-0,0614	0,9404**
		(0,5223)	(10,0330;12,0804)	(0,4518)	(0,0550;1,8259)
Estar Priv.	Saúde, Bem-	-0,4981	0,6077	-0,7941**	0,4520
		(0,4327)	(-0,2405;1,4559)	(0,3808)	(-0,2943;1,1983)
Neg. Dir Priv.	Ciê. Soc.,	0,9069**	2,4766***	-0,1191	0,8877**
		(0,4290)	(1,6357;3,3176)	(0,3787)	(0,1455;1,6300)
Comp. Priv.	Ciê., Mat.	0,1975	1,2184***	-0,9280**	0,3953
		(0,4573)	(0,3220;2,1147)	(0,4070)	(-0,4024;1,1931)
	Educ. Priv.	0,1405	1,1509**	-0,1706	0,8432**
		(0,4488)	(0,2713;2,0305)	(0,3964)	(0,0663;1,6201)
Const. Priv.	Eng., Prod.	-0,0913	0,9128*	-0,6748	0,5092
		(0,4938)	(-0,0552;1,8807)	(0,4286)	(-0,3308;1,3493)
Vagas Fortaleza	Oferta de	-0,0573***	0,9443***	0,0301***	1,0305***
		(0,0101)	(0,9244;0,9642)	(0,0096)	(1,0117;1,0493)
Vagas RMF	Oferta de	0,0393*	1,0401***	0,0225	1,0227***
		(0,0223)	(0,9963;1,0839)	(0,0213)	(0,9809;1,0646)
Vagas Interior	Oferta de	0,0235**	1,0238***	-0,0340***	0,9666***
		(0,0118)	(1,0007;1,0469)	(0,0109)	(0,9452;0,9880)
	RMF	-0,0229	0,9773***	-0,9177***	0,3994***
		(0,1437)	(0,6958;1,2589)	(0,1329)	(0,1390;0,6599)
	Prim. Emp.	-0,5163***	0,5968***	-0,2093***	0,8112***
		(0,0838)	(0,4325;0,7610)	(0,0659)	(0,6820;0,9403)
	Emp. Temp.	-0,2421	0,7850	-0,1345	0,8741*
		(0,6043)	(-0,3993;1,9693)	(0,4946)	(-0,0952;1,8434)
Empr. Atual	Exper.	0,1794***	1,1965***	-0,1701***	0,8436***
		(0,0260)	(1,1456;1,2475)	(0,0235)	(0,7975;0,8896)
	IMR	3,5117***	33,5042***	-2,8273***	0,0592
		(0,7068)	(32,1189;34,8894)	(0,5995)	(-1,1158;1,2341)
	Constante	-3,4220***	0,0326	4,1067***	60,7470***
		(0,7023)	(-1,3439;1,4092)	(0,6012)	(59,5687;61,9252)
	Observações		8,981		8,981
Likelihood	Log		-5,001		-5,991
Crit.	Akaike Inf.		10,063		12,043

Fonte: Elaboração dos Autores

Nota: *p<0,1; **p<0,05; ***p<0,01

No modelo dos determinantes dos desvios salariais, as variáveis que apresentaram ser estatisticamente significantes indicam as características individuais, educação e emprego atual.

Dentre as características individuais, quanto maior a idade do indivíduo menor a probabilidade deste auferir um salário abaixo do salário médio do seu curso de formação. Isto, possivelmente, devido ao acúmulo de experiência no cargo que ocupa acaba por se refletir sobre os salários. Se o indivíduo for homem reduz a probabilidade de este receber um salário abaixo da média correspondente ao seu curso de formação. Em termos percentuais, há uma menor chance de 26,6% dos homens receberem salários abaixo da média em relação às mulheres.

Com relação aos fatores educacionais, o fato de o indivíduo ter cursado seu Ensino Médio em escola pública diminui as chances de receberem salários abaixo da média do seu curso de graduação.³

Considerando o Ensino Superior Público, um indivíduo graduado em cursos nas áreas de Saúde e Bem-Estar, Ciências Sociais, Negócios e Direito, Ciências, Matemática e Computação e Educação possuem menor probabilidade de auferir salários abaixo da média salarial do seu curso de graduação. O mesmo resultado ocorre para os cursos nas áreas de Saúde e Bem-Estar e Ciências, Matemática e Computação para o Ensino Superior Privado. Esse resultado enfatiza que estas áreas de formação superior, que envolve Saúde, Educação e Tecnologia são as que possuem maior demanda por profissionais, refletindo em salários maiores.

Quanto aos fatores de expansão do ensino superior, um aumento no número de vagas implicará em uma queda na probabilidade de o indivíduo receber menores salários, quando essas vagas são direcionadas para o interior, onde a oferta de mão-de-obra qualificada é baixa. Em termos de chances, o aumento no número de vagas no interior diminui em 3,34% as chances de o indivíduo receber salário abaixo da média. Já se essas vagas se transferirem para a capital, há uma maior tendência de os indivíduos graduados receberem salários abaixo da média do seu curso de graduação. Em termos de chances percentuais, há um aumento de 3%. Isto porque, na capital há uma maior oferta de mão-de-obra com nível superior, o que leva a uma maior concorrência entre os indivíduos induzindo-os a aceitarem um salário abaixo da média.

De um modo geral, a expansão do Ensino Superior no interior aumenta a probabilidade de ser *overeducation* e diminui a probabilidade de auferir salários abaixo da média. Esse resultado pode ser explicado segundo Frenette (2004) em que a não existência da vaga correspondente a formação pode significar que mais graduados sejam forçados a aceitar empregos de nível mais baixo para evitar o desemprego, porém apresentando salários maiores. Ou seja, outros empregos não correspondentes ao nível de escolaridade acabam sendo mais atrativos. Esse comportamento se inverte, ao analisar o município de Fortaleza, onde observa-se uma menor tendência ao *overeducation* e com possibilidades de auferir salários abaixo da média do curso de formação.

As características relacionadas ao emprego, indica que se o indivíduo estiver em seu primeiro emprego, há uma redução nas chances de 19% de obter salários abaixo da média do seu curso. Essa probabilidade, também diminui com a experiência no emprego atual e com o fato do emprego se localizar na Região Metropolitana de Fortaleza.

³ A princípio, este resultado não era esperado. Acredita-se que este tenha sofrido uma influência da equação de participação no mercado de trabalho formal, uma vez que, esta foi utilizada para estimação da variável de *mills*. Os autores estimaram a equação dos determinantes sem a variável de *mills* e esta mostrou-se positiva e significativa, porém optou-se pelos resultados da regressão com a variável *mills* por considerá-la importante na estimação do modelo. Para mais detalhes, entrar em contato com os autores.

6 CONSIDERAÇÕES FINAIS

O conhecimento dos fatores, principalmente acadêmicos, que levam à situação de *overeducation* e desvio salarial são importantes, uma vez que a criação e manutenção dos cursos universitários, bem como outras políticas de expansão do ensino superior, implicam em custos à sociedade.

Dito isso, o presente trabalho tem por objetivo salientar potenciais determinantes do fenômeno *overeducation*, o descasamento entre o nível de escolaridade dos trabalhadores com a escolaridade exigida na ocupação atual e o desvio salarial. Para tanto, a investigação desses determinantes se restringe ao mercado de trabalho cearense para o ano de 2015.

Dentre os determinantes propostos, estão características observáveis relacionadas ao indivíduo, a sua educação como por exemplo, a área do curso de graduação e expansão do ensino superior, por fim, características sobre emprego atual do graduado.

A incidência de *overeducation* revela que os indivíduos tem em média 32 anos de idade, grande parte são mulheres e declararam-se não brancos. Em termos educação, a maioria cursou o ensino médio em escolas privadas. No ensino superior, poucos indivíduos receberam bolsa universitária e as IES estão avaliadas em média com conceito 2.75. Além disso, os cursos de instituições públicas na área de Agricultura e Veterinária apresentaram maior incidência de *overeducation* e na área da Saúde e Bem-Estar Social a menor incidência. Já nas IES privadas a incidência de *overeducation* fica nos cursos na área de Serviços e de menor, a área da Saúde e Bem-Estar Social e Engenharia, Produção e Construção. Acrescenta-se a evidência de uma maior incidência de *overeducation* quando há uma expansão das vagas de nível superior fora da capital cearense. Em termos de localização, grande parte dos *overeducated* estão na RMF e grande parte dos vínculos não são primeiro emprego, nem emprego temporário e em média os trabalhadores ficam 3 anos em um vínculo empregatício.

No modelo de desvio salarial abaixo da média, os indivíduos tem em média 30 anos de idade, maior participação de mulheres e de não brancos. Destaca-se ainda que a maioria cursou o ensino médio em escolas privadas. No ensino superior, poucos indivíduos receberam bolsa universitária e as IES estão avaliadas em média com conceito 2.75. Além disso, os cursos de instituições públicas na área de Serviços apresentaram maior participação de trabalhadores com desvio salarial abaixo da média. Já nas IES privadas os formados em Agricultura e Veterinária apresentaram essa característica. Quando há uma expansão das vagas de nível superior no interior cearense grande parte acabam por receber um salário abaixo da média. Em termos de localização, a maioria dos que recebem salário abaixo da média estão na RMF e grande parte dos vínculos não são primeiro emprego, nem emprego temporário e em média os trabalhadores ficam 3 anos em um vínculo empregatício.

Com relação aos potenciais determinantes do *overeducation*, as evidências empíricas revelam que o aumento da idade do formado, o fato dele ter cursado Ensino Médio em escola pública e o ensino superior em instituições públicas nas áreas de Agricultura e Veterinária, Serviços, Ciências Sociais, Negócios e Direito e Ciências, Matemática e Computação e Humanidades e Artes tendem a ter maior probabilidade dele ser *overeducated*. Caso a instituição seja privada e os cursos das áreas de Serviços e Ciências Sociais, Negócios e Direito também aumentam a probabilidade de o formado encontrar-se em uma situação de *overeducation* no mercado de trabalho cearense. Este resultado se aplica a medida que o formado adquire mais experiência no emprego atual.

Destaca-se ainda o papel da expansão do ensino superior ocorrida nos últimos anos. As evidências apontam para um possível efeito positivo da expansão de vagas no ensino superior sobre a probabilidade do formado está em uma situação de *overeducation*, quando essas vagas se direcionam para as regiões fora da capital Fortaleza. Aqui tem-se mais uma

evidência do efeito da interiorização do Ensino Superior, em que regiões de baixo dinamismo econômico e, portanto, mercado de trabalho mais restrito acabem por gerar um excesso de oferta de trabalhadores, de modo que para muitos, a saída para conseguir um emprego é ocupar cargos com funções não condizentes com sua instrução.

Já o fato do formado ter sido contemplado durante sua graduação com qualquer bolsa de estágio, extensão, monitoria ou pesquisa, ou ter cursado sua graduação em uma instituição de boa qualidade avaliado pelo IGC ou se sua área de formação corresponder a área de Saúde e Bem-estar Social do setor público ou estiver em seu primeiro emprego, diminui as chances dele ser *overeducated*.

Enfatiza-se ainda que a expansão do ensino superior na capital é um fator que pode influenciar o desvio salarial abaixo da média de salário do curso de formação. As áreas de Saúde e Bem-Estar, Ciências Sociais, Negócios e Direito, Ciências, Matemática e Computação e Educação ofertadas por IES pública e as áreas de Saúde e Bem-Estar e Ciências, Matemática e Computação ofertadas pelas IES privadas induzem a uma menor probabilidade dos formados de auferir salários abaixo da média salarial do seu curso de graduação.

Dadas as evidências encontradas, considerações políticas destinadas a favorecer a transição escola-trabalho de graduados que entram no mercado de trabalho devem ser adotadas. Em linhas gerais, enfatiza-se que não é somente a expansão desenfreada do ensino superior a solução para elevar o bem-estar social, mas a sua qualidade e o seu ajustamento com o mercado de trabalho local. Em termos de limitação de pesquisa, a falta de informações sobre *background* e composição familiar que podem afetar o status e a escolha do emprego atual, podem afetar alguns resultados. Além disso, sugere-se expandir a presente pesquisa para todos os municípios brasileiros, principalmente aqueles que receberam novos campi universitários devido a expansão do ensino superior.

REFERÊNCIAS

ANNEGUES, A. C. et al. Overeducation e Área de formação: evidências para os egressos da ufpb. In: 46º ENCONTRO NACIONAL DE ECONOMIA - ANPEC. Anais do 46º Encontro Nacional de Economia. Rio de Janeiro, 2018.

CAVALCANTI, M. F. A.; CAMPOS, F. M.; SILVEIRA NETO, R. d. M. Mismatch nos Mercados de Trabalho Regionais Brasileiros: O Que Explica as Diferenças Regionais? Encontro Regional de Economia, 2010.

DIAZ, M. D. M.; MACHADO, L. Overeducation e undereducation no Brasil: incidência e retornos. Estudos Econômicos, scielo, v. 38, p. 431 – 460, 09 2008.

DOLTON, P.; VIGNOLES, A. The incidence and effects of overeducation in the UK graduate labour market. Economics of education review, Elsevier, v. 19, n. 2, p. 179–198, 2000.

DOLTON, P. J.; SILLES, M. A. The effects of over-education on earnings in the graduate labour market. Economics of Education Review, Elsevier, v. 27, n. 2, p. 125–139, 2008.

DUNCAN, G. J.; HOFFMAN, S. D. The incidence and wage effects of overeducation. Economics of Education Review, Elsevier, v. 1, n. 1, p. 75–86, 1981.

FREEMAN, R. The Overeducated American. New York: Academic Press, 1976.

FRENETTE, M. The overqualified canadian graduate: the role of the academic program in the incidence, persistence, and economic returns to overqualification. *Economics of Education Review*, v. 23, n. 1, p. 29 – 45, 2004.

GROOT, W. The incidence of, and returns to overeducation in the uk. *Applied Economics*, Routledge, v. 28, n. 10, p. 1345–1350, 1996.

GROOT, W.; DEN BRINK, H. M. va. Overeducation in the labor market: a meta-analysis. *Economics of Education Review*, v. 19, n. 2, p. 149 – 158, 2000.

HARTOG, J. Over-education and earnings: where are we, where should we go? *Economics of Education Review*, v. 19, n. 2, p. 131 – 147, 2000.

HECKMAN, J. J. Sample selection bias as a specification error. *Econometrica*, [Wiley, Econometric Society], v. 47, n. 1, p. 153–161, 1979.

LEUVEN, E.; OOSTERBEEK, H. Overeducation and mismatch in the labor market. *Handbook of the Economics of Education*, v. 4, p. 283–326, 2011.

MARIONI, L. Overeducation in the labour market: Evidence from brazil. In: ROYAL ECONOMIC SOCIETY. Conference 2018 Royal Economic Society Annual Conference. Brighton, 2018.

MCGUINNESS, S. Overeducation in the labour market. *Journal of Economic Surveys*, Wiley Online Library, v. 20, p. 387–418, 02 2006.

MORANO, C. P. The determinants of overeducation: Evidence from the Italian labour market. In: ARACIL, A. G.; GÓMEZ, I. N. (Ed.). *Investigaciones de Economía de la Educación 9*. [S.l.]: Asociación de Economía de la Educación, 2014, (*Investigaciones de Economía de la Educación* volume 9, v. 9). cap. 35, p. 681–698.

NIETO, S.; RAMOS, R. Overeducation, skills and wage penalty: Evidence for spain using piac data. *Social Indicators Research*, v. 134, n. 1, p. 219–236, Oct 2017.

PRITCHETT, L. Where has all the education gone? *The World Bank Economic Review*, Oxford University Press, v. 15, n. 3, p. 367–391, 2001.

QUINN, M. A.; RUBB, S. Mexico's labor market: The importance of education-occupation matching on wages and productivity in developing countries. *Economics of Education Review*, v. 25, n. 2, p. 147 – 156, 2006.

REIMER, D.; NOELKE, C.; KUCEL, A. Labor market effects of field of study in comparative perspective: An analysis of 22 European Countries. *International Journal of Comparative Sociology*, v. 49, n. 4-5, p. 233–256, 2008.

REIS, M. C. Educational mismatch and labor earnings in brazil. *International Journal of Manpower*, v. 38, n. 2, p. 180–197, 2017.

ROBST, J. College quality and overeducation. *Economics of Education Review*, v. 14, n. 3, p. 221 – 228, 1995.

ROCHA, R. H. et al. A relação entre o ensino superior público e privado e a renda e emprego nos municípios brasileiros. *Pesquisa e Planejamento Econômico*, v. 47, n. 3, p. 39–69, 2017.

TSANG, M. C.; LEVIN, H. M. The economics of overeducation. *Economics of Education Review*, v. 4, n. 2, p. 93 – 104, 1985.

VERHAEST, D.; OMEY, E. Discriminating between alternative measures of over-education. *Applied Economics*, Taylor & Francis, v. 38, n. 18, p. 2113–2120, 2006.

WOOLDRIDGE, J. *Introductory Econometrics: A Modern Approach (with Economic Applications, Data Sets, Student Solutions Manual Printed Access Card)*. 4. ed. [S.l.]: South-Western College Pub, 2008.

AN UNINTENDED CONSEQUENCE OF VOCATIONAL SCHOOLS' CONSTRUCTION: EVIDENCE FROM CEARÁ

Rafael B. Barbosa*

Alessandra Benevides**

Francisca Zilânia Mariano***

RESUMO

Este artigo analisa o efeito dos pares sobre os resultados educacionais de estudantes do ensino médio. Para identificar a relação causal, utiliza-se uma variação exógena na qualidade dos estudantes das escolas regulares no Ceará, decorrente da expansão das escolas profissionalizantes. Escolas profissionalizantes realizam um processo seletivo dentre os estudantes do primário e com isso afetam a qualidade dos estudantes das escolas regulares. Os resultados sugerem que reduzir a proporção de pares com elevado desempenho afeta negativamente as notas em testes padronizados de matemática e português dos estudantes das escolas regulares. Além disso, a qualidade dos pares impacta sobre a evasão escolar e a taxa de repetência durante o ensino médio. Tais resultados parecem ser explicados pelo aumento da distração dos estudantes, diminuindo seu foco nos estudos.

Palavras-Chave: Efeito dos pares, efeito de equilíbrio geral, Ensino Médio

ABSTRACT

This paper investigates the effect of reducing peer quality on high-school students' outcomes. To identify the causal effect, we exploit an exogenous variation in peer composition caused by a large program of selective public schools in Brazil. Results suggest that students exposed to an increase in the proportion of low-achievers diminish test scores, both math and language, and raise the rate of retention during secondary education. The channels that explain the underlying findings are associated with an increase in students' attention diversion and a reduction of persistence in learning. We do not find evidence that the teacher labor market and teacher pedagogical practices are affected by the change students' composition.

Keywords: Peer effect, General equilibrium effect, Secondary education

JEL: I2, I28, I24

ÁREA 4: Setor Público

1 INTRODUCTION

Social interactions are a fundamental determinant of several economic outcomes¹. However, the effect of peer in education, notably on secondary education, is fiercely debated. Peers can influence many aspects of an individual's economic decisions, especially during

* Universidade Federal do Ceará. Email: rafael.barbosa@ufc.br/ tel.: (85)99930-1938.

** Universidade Federal do Ceará.

*** Universidade Federal do Ceará.

¹ There are many evidences about peer effect in different areas as health (Fortin; Yazbeck, 2015), crime (Billings; Deming; Ross, 2016), job productivity (Georgeanas *et al.*, 2015, Falk; Ichino, 2006), and education (Sacerdote, 2014).

adolescence, when the student is subject to several behavioral biases. Furthermore, it is not clear whether high-achievers peers positively affect students, neither the mechanisms that could explain the potential effects. These debates are, in part, related to econometric identification's problems as self-selecting (sorting), endogenous, and reflection bias (Manski, 1993, Angrist, 1994). Recent literature has followed a series of new identification strategies to overcome these limitations.

In this paper, we consider a plausible quasi-natural experiment that changes the peer composition for some secondary schools in Brazil. Specifically, using a difference-in-difference empirical strategy, we find that increasing the proportion of low-achievers affects the students' test scores and other educational outcomes, as the rate of retention and drop-out. To measure exposure to low-achievers peers, we exploit a large state's program of selective public schools' construction in the Ceará, a Northeast state of Brazil. Since 2008, the state of Ceará has expanded the number of vocational schools (*Escolas Profissionalizantes*) relative to regular ones². Unlike other secondary schools in Brazil, vocational schools in Ceará execute an admission process that selects high-achievers from middle education (Ensino Fundamental II). This selection process separates high and low-achievers during secondary education³ in municipalities that receive vocational schools.

The construction of vocational schools significantly reduces the proportion of high-achievers students attending regular education. We assume that the timing, location, and the share of students attending vocational schools are orthogonal to unobservable regular students' factors. We extensively test the validity of these assumptions. We use the proportion of students attending vocational schools as a measure of exposure to low-achievers peers for regular students.

Theoretically, the impact of change the peer composition is ambiguous (Duflo, Dupas, and Kremer, 2011; Lazear, 2001). Unfortunately, empirical pieces of evidence are also controversial. Lavy, Paserman, and Schlosser (2012) estimate that the proportion of repeaters harms the academic achievements of peers. Carrell, Hoekstra, and Kuka (2018) find similar results, indicating that the presence of disruptive students in elementary schools negatively affects long-run outcomes, as earnings and college attendance. However, Duflo, Dupas, and Kremer (2011) point out that homogeneous classrooms raise students' test scores, even with a large number of low-achieving students. In this case, the mechanism is explained by teachers' adjustment of pedagogical practices. We expect to shed light on this controversy providing robust empirical evidence of the mechanisms which peers potentially affect other students.

This paper has two main contributions. First, it estimates the effect of reducing peer quality on regular students' outcomes exploiting a plausible exogenous variation in the share of students enrolled in vocational schools during secondary education. We consider as the main outcome variable the student's test scores in math and language at the end of secondary education (12th grade). Moreover, we also assess the effect on long-run outcomes as the rate of student drop-out and retention using a school-level version of the empirical strategy. Both outcomes are relevant to human capital accumulation and may affect the economic opportunities of the student's life.

² In Brazil, regular schools teach the standard high school curriculum and generally have four or five class hours per day.

³ In addition, vocational schools are preferred by parents and students in comparison to regular ones. Parents prefer vocational schools for two reasons. First, vocational school is full-time, contrary to the regulars, which is part-time. Therefore, parents can leave the students under the school's supervision during all workday. Second, vocational schools provide vocational skills, increasing the probability of students enters the labor market after the secondary. In some municipalities of Ceará, the supply of tertiary education is limited, then attending vocational schools can improve the economic opportunity.

Second, we extensively attempt to understand the mechanisms that explain underlying findings, analogously to Lavy, Paserman, and Schlosser (2012). We consider several potential channels as student engagement, interest in learning, student well-being, time allocation, personality traits, and teacher's pedagogical practices. These channels are related to main theories that explain how peers affect her colleagues.

The results suggest that the share of students attending vocational schools negatively affects the academic achievements of regular students. The estimate effect size is large in comparison with other educational policies. The magnitude of the impact in math is higher than in language test scores; however, the difference is small.

We also test the presence of a heterogeneous effect. First, we find that girls are more affected by peers' change in composition than boys in math test scores. Second, students residents in small municipalities are less impacted by the reduction of peer quality than students living in big cities. Third, the results also suggest the peer effect is highly nonlinear. Students with a low performance before secondary education (9th grade) are more negatively affected by their peers. Interestingly, we find a positive effect of the change in peer composition in high-achievers' students.

Furthermore, a concentration of low-achieving students raises the rate of drop-out and retention during secondary education. The impact on drop-out is more prominent in the small municipalities. This paper also contributes to evidence about the effect of peers on long-run outcomes as reported by Carrell, Hoekstra, and Kuka (2018), Gould, Lavy, and Paserman (2009), Bifulco, Fletcher, and Ross (2011), Bifulco et al (2014), Anelli and Peri (2017). Results for both short and long-run outcomes are submitted several robustness' checks, and we conclude that our estimates are unlikely to be biased.

We then move to understand the channels that explain the underlying estimates, examining teacher and student surveys. Results suggest that girls exposed to a high proportion of low-achievers reduce the expectation about success in the future and increase the time spending on the internet. Boys are not affected by the change in peer composition. These conclusions support the evidence of large peer effects on girls. We do not find evidence that changes in peer composition impact teachers' pedagogical practices. We confirm our results using a teacher survey about her pedagogical practices and student perception about teachers.

At last, we verify the effect of reducing peer quality on students' personality traits. We measure the personality traits using a Big-Five instrument, called SENNA Instrument, implemented by Instituto Ayrton Senna, a Brazilian civil society organization focusing on citizen agency and public-service delivery. The SENNA Instrument measures the five constructs of the Big-Five inventory⁴ and the facets that make up each construct. We estimate the effect of vocational school constructions against the facets of Conscientiousness and Openness, both associated with educational achievement. Our estimates suggest that vocational school This paper contributes to three strands of literature. First, we contribute the literature about the importance of peer composition for student outcomes as Lazear (2001), Duflo et al (2011), Lavy et al (2012), Imberman et al (2012), Carrell et al (2018), Shiltz et al (2019), Bossavie (2020), and others. By using a quasi-experimental approach, we cope with the main econometric challenges in peer effect estimation. Also, our measure of the share of low-achievers does not depend on specific students' aspects, as repeaters (Lavy et al (2012)) or disrupt contemporaneous behavior and learning (Carrell et al (2010, 2018)) that may cause difficulty in the effect interpretation. Second, a few papers study how education policies might have general equilibrium effects on student outcomes (Duflo (2004), Bianchi (2020), Gilraine

⁴ Big Five Inventory is a five-factor model of personality traits formed by: Conscientiousness, Openness, extraversion, agreeableness, and neuroticism. For more details see Almond et al () and the Appendix A.#.

et al (2018)). This paper complements this literature by showing how the expansions of vocational schools have unintended spillover effects on students of different schools. Third, this paper sheds light on the channels through the social interactions impact students. A few papers study these mechanisms, especially in secondary schools (Lavy et al (2012)). We add to these studies by evidencing that the main channels of peer effect are related to diverting student attention, expectations about the future, and time allocation.

2 BACKGROUNDS

2.1 Vocational schools in Ceará, Brazil

The state of Ceará is located in the Northeast of Brazil, one of the poorest regions in the country. Ceará's per-capita GDP is nearly USD 5.500, which is smaller than the average per-capita GDP in Brazil (close to USD 9,800). The state's population is approximately 9 million, and the Human Development Index (HDI) is 0.68, close to Brazil, 0.695.

Secondary education in Ceará is segregated into three kinds of schools: regular, vocational, and indigenous education.⁶ Regular schools are part-time education and teach only the standard secondary nationwide curriculum, providing skills to students that aim to apply to higher education at the end of secondary. Indigenous schools teach only a specific indigenous curriculum and represent a small share of secondary school in Ceará, nearly 1%.

Since 2008, the government of Ceará has expanded the number of vocational schools. In 2017, there existed 120 vocational schools in Ceará, representing nearly 17% of the secondary education in the state. Vocational schools have remarkable differences comparing with regular ones. First, vocational schools have a full-time program and two different curriculums. Half of the day class, students learn a standard nationwide curriculum, similar to regular counterparts. In the other half time, students learn practical knowledge and occupation-specific skills that directly map into entering a particular occupation in the labor market. Second, and more relevant for our purposes, vocational schools in Ceará perform an admission process that cherry-picks the high-achievers. This admission process reduces the proportion of high-achievers in regular schools during secondary education. In general, vocational students have better socioeconomic characteristics, higher levels of cognitive and non-cognitive aspects than regular ones.

The admission process differentiates vocational from regular students in at least two aspects. First, vocational students show intrinsic motivation to study in these schools because they have to apply to compete for a slot. Demotivated students do not apply and do not participate in the selection process. Second, the admission process selects students based on high-stakes performances over the last three academic years. Therefore, vocational students have higher grades and also sustained their performance for a long time in comparison with other students. This aspect is associated with persistence and grit (Duckworth et al., (2009)).

Table 1 summarizes the differences between regular and vocational students. We consider test scores⁷ and many other demographics and socioeconomic aspects as the proportion of female, age, race, mother's education, and the share of students that at least one family member a conditional cash transfer (CCT) program. The average scores of vocational students are 27.32 and 13.21 points higher than regular students in math and language, respectively.

⁵ To put in perspective, Ceará's GDP is similar to countries like Nicaragua, Mozambique, Albania. Moreover, the Ceará population is similar to Paraguay, a Latin-America country.

⁶ There exist a fourth school category since 2018, the full-time schools. However, our data span until 2014.

⁷ We measure cognitive skills using administrative data provide by Ceará's Department of Education (SEDUC). More details are described in section #. The distribution of these tests has a mean 250 and a standard deviation of 50 points.

Regular students are represented by a smaller proportion of girls and white students. In turn, vocational students are younger, their mothers are more educated, and they live in families that have a lower proportion of parents attending the cash-transfer program. These results suggest that vocational students have high cognitive skills and better socioeconomic conditions than regular students.

During the implementation process of vocational schools, we observe variation in the exposure of students over time and geographical areas. The government smoothly scaled up the program across years and municipalities. Figure 1 shows the municipality evolving of vocational schools along the time between 2010 to 2017.

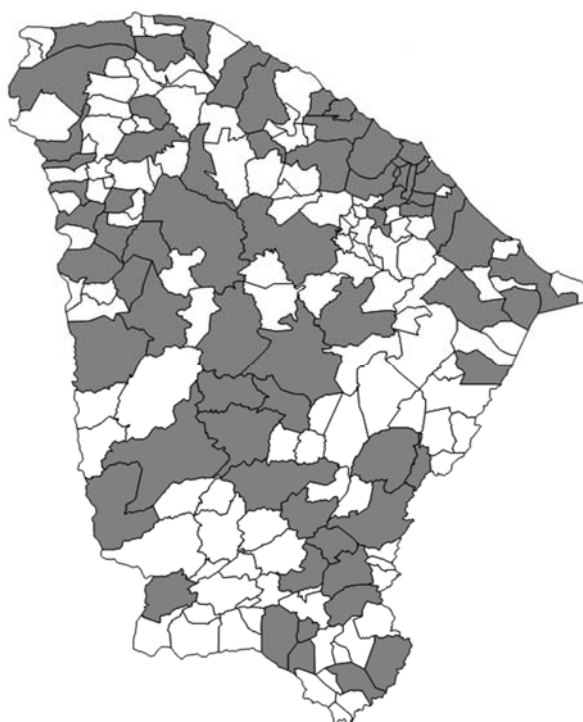
Vocational school is usually preferred by parents and students. The Ceará's state has a small offer of public higher education. Students that attend regular education and do not apply to public universities have few opportunities to enter the labor market. Vocational schools aimed to raise the opportunities for students by providing them occupation-specific skills, allowing enter the labor market just after the secondary.

3 DATA AND DESCRIPTIVE STATISTICS

3.1 Data Sources

This paper requires considerable data sources. First, we create the share of students attending vocational schools in each municipality of Ceará's state from the annual School Census, a survey of every school in Brazil, conducted by the Ministry of Education. We consider only students that attended public middle education before enrolling in high-schools, excluding students from private schools. The School Census data also allows measuring some school quality indicators, as the proportion of teachers with higher education and the average number of students per classroom.

Figure 1: Vocational School Construction in Ceará



Notes: Figure 1 presents the variation in time and geographic location of the expansion of vocational schools in Ceará's state, Brazil. The first vocational school was introduced in 2008 and in 2017 there exists 120 schools in Ceará, representing 17% of all secondary schools.

To measure test scores in math and language, we consider the administrative data from SPAECE⁸, a state test applied by Ceará's Department of Education (SEDUC). The SPAECE data span from 2008 to 2014 and include individual test scores for students in the 9th grade of the middle and 3rd grade of high-school (9th and 12th grades, respectively). The data contains 550.867 students at the end of secondary (12th grade), where nearly 93% of students attend regular schools.

The Ceará's Department of Education also applies, during SPAECE, a contextual survey to students that allows construct student demographic and socioeconomic characteristics. In some cases, we were not able to match student test scores to student's survey data because of misreporting. We test different specifications of empirical strategy to cope with this missing information.

The contextual survey also provides information about student engagement, perception about teacher practices, student well-being, and time allocation. We use this information to understand the potential mechanisms of the peer effect. SPAECE also applies a survey for teachers. We use the teacher survey to understand whether their pedagogical practices are associated with vocational school construction⁹.

We verify if the construction of the vocational school is associated with the student's personality traits. To measure the personality traits, we consider an instrument applied on census-basis to Ceará students only in 2015. The instrument, called the SENNA instrument, was conducted by the Instituto Ayrton Senna, a Brazilian civil society organization. This instrument allows capturing the Big-Five personality traits and its facets. We focus our estimates on the Conscientiousness and the Openness' facets because are more related to students' performance, according to literature. The facets of Conscientiousness are Determination, Focus, Organization, and Persistence. In turn, the facets of Openness are Curiosity to Learn, Creative Imagination, and Artistic Interest. Our results about personality traits should be seen as preliminary, however, since this relies on data for a single year.

Finally, we consider as long-run outcomes the rate of school drop-out and the rate of retention during high school. These variables are constructed from the Ministry of Education at the school level. Both variables represent the percentual of students that dropped out or repeated during one of the three years of secondary schools.

3.2 Descriptive Statistics

Based on the admission process, we posit that students attending vocational schools have better performance in comparison with regular students. Figure 2 displays the distribution of test scores in math and language at the ending of grade 9, before they enter in high-school. The scores are standardized to have mean zero and standard deviation one. As suggested by this graph, vocational students outperform regular students in math and language test scores. This evidences that regular students have low-abilities compared with vocational students before they attend vocational schools.

We are interested in the causal effect of high achieving peers on students left behind in regular education. Hence, the treatment group is composed of regular students exposed to

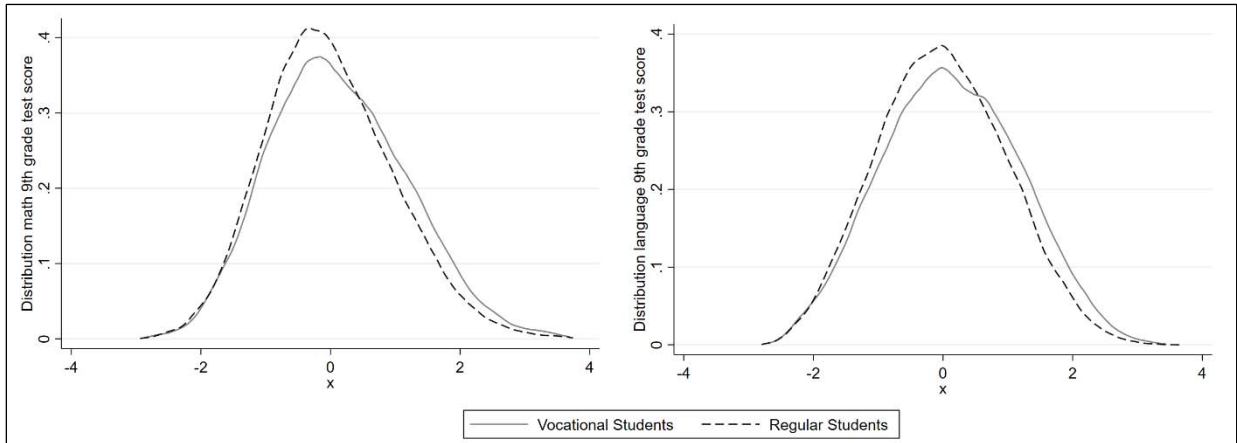
⁸ Sistema Permanente de Avaliação da Educação Básica do Ceará (SPAECE).

⁹ We do not have information about these surveys for all cohorts. Many questions change along the cohorts and we consider items with large size possible. Details of data, see Appendix A#

vocational schools. Table A2, reported in appendix, compares treated and untreated students, i.e. the regular students that were not exposed to vocational school construction.

Treated and untreated students are very similar. There is no statistical difference among scores in math and language at the end of middle education, grade 9. Treated students outperform untreated students on language scores at the ending of secondary. Untreated students are represented by a large share of black and brown and are younger than treatment group. In turn, the treated students are less poor, and their mothers have more years of instruction.

Figure 2: **Distribution of math and language test scores in 9th grades**



Notes: Figure 2 presents the distribution of math and language test scores in 9th grades for students that in the next year will attend regular (black) and vocational (gray) schools. It shows that vocational students have better performance in both subjects before enroll in secondary education.

In short, treated and untreated students has similar characteristics and we expect that the exposure to vocational schools is the only factor that differentiates short and long-run educational outcomes between the two groups.

Table 1: **Descriptive statistics comparing vocational and regular students**

Variables	Vocational Students		Regular Students		Difference
	Average	SD	Average	SD	
Test Scores Math 9 ^o grade	255.74	47.94	228.42	45.71	27.32***
Test Scores Portuguese 9 ^o grade	249.88	48.61	236.67	44.67	13.21***
Girls	0.57	0.49	0.55	0.49	0.02***
Black	0.11	0.31	0.14	0.35	-0.03***
Brown	0.11	0.31	0.12	0.33	-0.01
White	0.17	0.38	0.08	0.28	0.09***
Age	14.59	1.60	15.34	2.80	-0.75***
Mother education	2.94	1.07	2.85	1.07	0.09***
Attend Bolsa Familia	0.72	0.44	0.74	0.43	-0.02**

Note: Table 1 presents the descriptive statistics of vocational and regular students' factors. The last column shows the difference about the variables' averages. The stars represent the rejection of the null hypothesis of equal mean based on a test t. Significance levels: 1% ***, 5% **, 10% *.

4 EMPIRICAL STRATEGY

To overcome the usual problems of selection and sorting associated with the estimation of peer effects, we rely on exogenous variation in timing and location of vocational school construction. We also access the intensity of vocational school construction's exposure by using the share of students that, in some municipality and cohort, attend vocational school.

Our benchmark specification is the following

$$y_{ismt} = \beta + \gamma Share_voc_{mt} + \delta' X_{ismt} + \alpha' X_{smt} + \tau_t + \pi_s + \varepsilon_{ismt} \quad (1)$$

Where y_{ismt} is the achievement for student i , on the school s , in municipality m , on year t ; $Share_voc_{mt}$ is the share of students attending vocational schools in municipality m in year t . π_s is a school effect, τ_t is a time effect, X_{ismt} is a vector of students controls that includes gender, race, age, mother's education, an indicator that a family is a recipient of a conditional cash transfer program (CCT), 9th grades test scores in math and language, polynomials third-order of these test scores, an indicator that student usage public transportation to go to school. X_{smt} is a vector of school covariates that includes the proportion of teachers with tertiary education and average class size. When presenting our estimates, we show different combinations of these covariates. ε_{ismt} is the error term. Standard errors are clustered at the school level.

Our parameter of interest is γ , which measures the effect of the regular students being exposed to an increase of the proportion of high-achievers attending vocational schools. We assume that the rise of the share of students attending vocational schools implies a reduction of the quality of peers in the regular schools, measured by the proportion of low-achiever's students.

To control for potential confounding factors, we include in all specifications school and time fixed effects. However, one may be concerned that there are time-varying unobserved factors that are also correlated with the proportion of low-achieving students at school level. Therefore, we also estimate a model adding a full set of school-specific linear time trends to (1).

We also consider a school-level version of (1) that is used to estimate the effect of increase the share of low achieving students on the rate of student drop-out and retention. These variables have high opportunity costs for the students suggesting that the peer quality can have a persistent effect.

We estimate the following equation

$$y_{smt} = \beta_0 + \mu Share_voc_{mt} + \vartheta' X_{smt} + \tau_t + \pi_s + \varepsilon_{smt} \quad (2)$$

Where y_{smt} is the proportion of repeaters or students that dropped out the regular school s in the year t in the municipality m . Important, we consider these variables for all high school grades (10, 11, and 12 grades). In the Online Appendix, we estimate the same specification for the three grades of high-school separately.

To control for school potential confounding variables, we include specific school controls, X_{smt} , and school fixed effect. X_{smt} contains the proportion of teachers with tertiary education and average class size. Therefore, we also consider a model that includes school interacted with time fixed effects to capture time-varying school-specific unobservable factors.

4.1 The validity of the identification strategy

The validity of identification strategy depends on two key assumptions: (1) time and location of vocational school construction are exogenous to regular students, (2) the share of students attending vocational schools is unrelated to regular students' unobservable factors.

The first assumption is not validated if the decision about the time and location where constructed a vocational school is related to students' unobservable characteristics. To assess this possibility, we regress a logit panel model where the dependent variable is the time and location of vocational schools against a vector of controls that includes average students age, proportion of girls, racial shares, and average previous test scores in language and math for 9th and 12th grades. The results do not suggest that the decision to construct a vocational school is reasonably associated with these factors¹⁰. The demographic factors are not significant and the estimates for average grades present contradictory implications. For example, the results indicate that vocational school constructions are correlated to municipalities with higher average language test scores and lower test scores in math.

In turn, the share of students attending vocational schools are unrelated with regular students' unobservable factors is plausible assumption because the decision to be enrolled in a vocational school is took before the secondary education. Thus, students accepted to enroll in vocational school do not previously know the quality of future peers.

A potential threat to this assumption is the students' capacity to anticipate the quality of future peers in high school based on the quality of current peers in middle schools. We consider this possibility unlikely for two reasons. First, students probably change the school during the transition from middle to secondary education because the number of middle schools is much higher than secondary education¹¹. Second, many students dropped out during the transition to middle to secondary education¹². Both reasons affect the capacity of predict the quality of the peers on secondary.

The contributions of Manski (1993) has evidenced the fundamental problem of selection into peer groups which can contaminate peer effect estimates. First, students may self-select themselves into peer groups based on certain unobserved factors, called correlated effects. Second, peers may influence each other simultaneously, known as reflection problem. Third, it is difficult to distinguish between peer effects due to peers' achievement, endogenous effects, and peer effects due to peers' background, contextual peer effects.

Our empirical strategy overcomes the first and second fundamental problems. The construction of the vocational school has a municipal level impact, i.e. which is not school-specific. This minimizes the self-selection process, especially the exposure of the regular students to high-achievers. In addition, the inclusion of school fixed effects accounts for the most obvious source of student sorting between schools.

¹⁰ The results of these estimates are not reported in this paper by concision; however, it is available under the authors' contact by emails.

¹¹ Particularly in Ceará, there existed 717 secondary schools and 4326 middle schools in 2017.

¹² The rate of school progression from middle to secondary education is 86%.

Table 2: Change in the students' characteristics

Panel A: All municipalities						
	(1)	(2)	(3)	(4)	(5)	(6)
	Girls	Black	Brown	Age	EJA	Urb
Treat	-0.00304 (0.00187)	-0.0016 (0.00117)	-0.0081 (0.0135)	0.021 (0.044)	0.00609* (0.00333)	-0.00349 (0.00539)
Observations	2,601,977	2,601,977	2,601,977	2,601,977	2,601,977	2,601,977
R-squared	0.006	0.014	0.202	0.294	0.774	0.37
Panel B: Small municipalities						
	(1)	(2)	(3)	(4)	(5)	(6)
	Girls	Black	Brown	Age	EJA	Urb
Treat	-0.00132 (0.00218)	-0.00145 (0.00142)	0.000377 (0.0151)	0.0191 (0.0544)	0.00664 (0.00414)	-0.00732 (0.0071)
Observations	1,782,016	1,782,016	1,782,016	1,782,016	1,782,016	1,782,016
R-squared	0.005	0.012	0.2	0.239	0.687	0.214
School FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y

Notes: Table 2 presents the impact of vocational school construction on municipal students' factors. It evidences that only the proportion of students using public transport is significantly affect by the vocational school program. Each estimation includes municipal and time fixed effects. Significance levels: 1% ***, 5% **, 10% *.

The last concern is the possibility of students migrate when vocational schools are constructed. If the future regular or vocational student migrates to a different municipality when the vocational school is constructed, then the results could be biased. To test this possibility, we estimate the effect of the vocational school construction on the municipal students' demographic and socioeconomic characteristics, like the proportion of girls, the racial shares, age, special educational programs¹³ (EJA), and the share of urban students. If the vocational school introduction produces a relevant migration process, we expect that municipal students' characteristics will be affected. Table 2 presents the estimates. All models include municipal and time fixed effects.

Panel A shows the estimates for all municipalities and Panel B consider only small municipalities¹⁴. We do not find evidence that vocational school construction affects the student demographic characteristics, suggesting an absence of migration caused by the presence of selective schools. In summary, the results corroborate the validity of the empirical strategy's assumptions.

5 RESULTS

5.1 Effects on test scores

Tables 3 and 4 report the effect of the proportion of vocational students on high school achievement of regular students in math and language, respectively. We transform SPAECE's test scores into standardized z-scores to facilitate the interpretation of the results. We consider six specifications in which the differences stem from the additional covariates. The presence of missing data in the sample reduces the sample size as more covariates are incorporated.

Column 1 presents the average treatment effect and standard deviation of the outcome variables for regular students considering only additive school and time fixed effects. This sample has nearly half a million students at the end of high-school, spanning from 2008 to 2017.

¹³ Older students can opt to be enrolled in special educational programs (EJA) that reduce the time of secondary education.

¹⁴ We refer to a small municipality if their population is lower than 250 thousand inhabitants. Only three municipalities have population higher than 250 thousand inhabitants: Fortaleza (State's capital), Caucaia, and Juazeiro do Norte.

The columns 2-5 include additional controls. Column 2 considers regular controls as gender, age, racial status, and an indicator of the use of public transport. Column 3 adds 9th test scores in math and language to control for previous skills. Column 4 adds as control an indicator for students' families that receive conditional cash-transfer (CCT) and mother's education. Finally, Column 5 includes the third-order polynomials for math and language 9th test scores. Column 6 considers the same specification of column 5, however, it adds the school-by-time fixed effects which controls for potential time-varying unobservable factors.

Results show that all estimates are negative and statistically significant. This suggests that the share of students attending vocational school reduce the performance of students from regular schools. The estimates do not change in magnitude in different specifications, except for Column 3, which includes previous students' test scores. This result can be associated with the sample reduction. However, in more restrictive samples, as Columns 4, 5, and 6 the magnitude of initial specifications is restored.

The average effect size is -0.38σ for language and -0.40σ for math considering three years of exposure to low-achievers peers. The impact of reducing peer quality affects mathematical and reading skills, suggesting that the global effect on student ability can be larger¹⁵. To put these estimates into perspective, considering that, in 2017, 17% of secondary students in Ceará are enrolled in a vocational school, then the average effect is near -0.06σ for math and language.

Table 3: Estimates of the proportion of vocational students on math achievement of regular school

Math	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.410*** (0.065)	-0.347*** (0.076)	-0.165* (0.093)	-0.419*** (0.131)	-0.444*** (0.128)	-0.498*** (0.049)
School fixed effects	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Regular control		Y	Y	Y	Y	Y
Prior test scores			Y	Y	Y	Y
Additional control				Y	Y	Y
Non Linearities					Y	Y
School-time fixed effects						Y
N. obs.	502.920	364.260	167.499	96.176	96.176	96.176

Notes: Table 3 reports the effect of share of vocational students on math test scores of regular students. The five specification change according to the number of covariates in each model. The specifications (1) to (5) contains school and time fixed effect. The column (6) presents results for the school-by-time fixed effects. The standard errors are estimated clustering by school. Significance levels: 1% ***, 5% **, 10% *.

These results evidence that peer quality is relevant to students' achievements. Specifically, our results indicate that peer quality matters to test scores at the end of secondary education in Brazil. Although the effect size is not large, we can compare it with other studies. Jackson (2014) analyzes the effect of teacher quality on high-school students in North Carolina State, US. The impact of a decrease by one standard deviation the teacher quality is 0.06, similar to our estimates. Then, to put in perspective, the effect of reducing peer quality is equivalent to

¹⁵ In Online Appendix, we present the estimation on the sum test scores. The point estimate is -0.79σ for our preferred specification similar to column 5 in both tables.

reducing teachers' value-added by one standard deviation, according to estimates from Jackson et al (2014).

Table 4: Estimates of the proportion of vocational students on language achievement of regular school

Language	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.490*** (0.061)	-0.379*** (0.066)	-0.258** (0.085)	-0.330** (0.118)	-0.337** (0.116)	-0.359*** (0.048)
School fixed effects	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Regular control		Y	Y	Y	Y	Y
Prior test scores			Y	Y	Y	Y
Additional control				Y	Y	Y
Non Linearities					Y	Y
School-time fixed effects						Y
N. obs.	502.920	364.260	167.499	96.176	96.176	96.176

Notes: Table 4 reports the effect of share of vocational students on math test scores of regular students. The five specification change according to the number of covariates in each model. The specifications (1) to (5) contains school and time fixed effect. The column (6) presents results for the school-by-time fixed effects. The standard errors are estimated clustering by school. Significance levels: 1% ***, 5% **, 10% *.

5.1.1 Heterogenous effects on test scores

There exists a fierce debate about the presence of non-linearity of peer effect. For example, the monotonicity model posits that the lower the quality of the peers the bigger is the negative effect on students. Sacerdote (2011) suggests that nonlinearities in the relationship between own and peer characteristics may help explain the heterogeneity evidence about peer effects. Several studies have found evidence of nonlinear peer effects, wherein lower ability students lose more from decreases in peer quality than do higher ability students (Hoxby and Weingarth (2005), Imberman et al. (2012), Burke and Sass (2013), and Fu and Mehta (2018)). We test the presence of non-linear effects on regular students due to the peer composition changes as a consequence of vocational school introduction.

We verify the effect of being exposed to large peer quality reduction on regular students considering the student's educational status in 9th grade. The Ceará's Department of Education classifies the students in the 9th grade in four levels according to performance in test scores: Very Critic, Critic, Intermediate, and Advanced¹⁶.

We estimate our preferred specification that includes all covariates similar to Column 5 and 6 in Table 4. To assess the heterogeneous effect, we interact the exposure measure, $Share_{voc}_{mt}$ with each of the student's prior educational levels, computed in 9th grade. Figure 3 depicts the effect on each student group level.

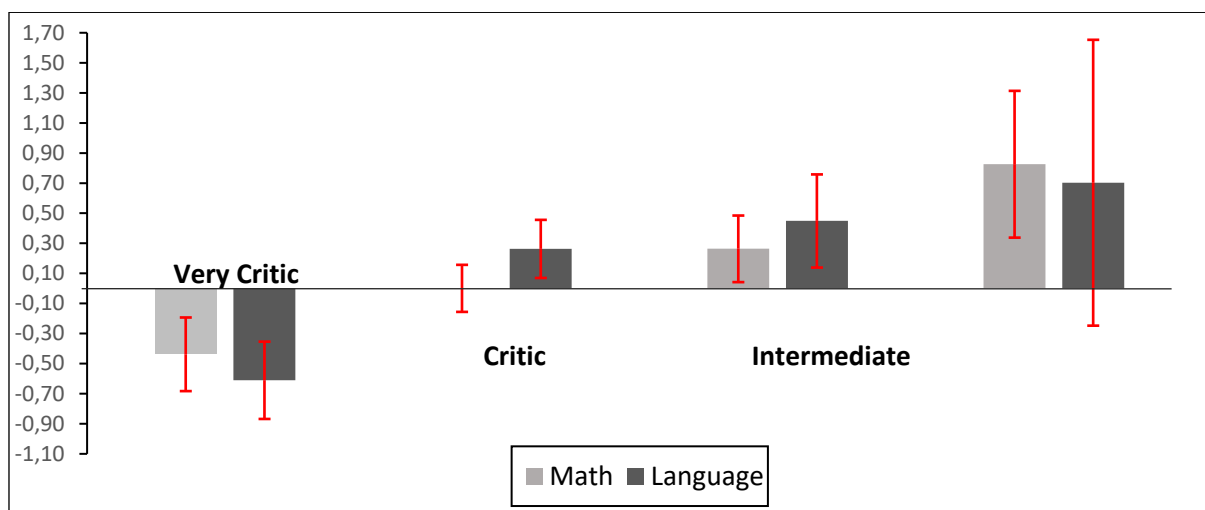
Students at a Very Critic level are strongly negative affected in both math and language. This suggests that the validity of the monotonicity model, i.e., worsen peers lower the test scores of their schoolmates. The magnitude of the effect decreases according to the

¹⁶ This approach is more suitable to testing nonlinear models using our data. The standard approach is based on quartile specifications as Imberman et al (2012) or Hoxby and Weingarth (2006). Our sample has missing values in 9th-grade, which prevents obtaining the reliable ranking of the student in 9th grade.

educational level raise, indicating the presence of weak monotonicity (IMBERMAN et al., 2012)).

Specifically for math, students at the Intermediate and Adequate levels benefit if more students attend a vocational school. This indicates the presence of the invidious model, which posits that a student's performance increases by having less able peers. Consistent with the previous literature, we find that changing peer composition has a nonlinear effect on their colleagues, in which low-achievers' students are more harmed and high-achievers are positively impacted.

Figure 3: Heterogeneous effect of reducing peer quality on test scores



Notes: Figure 3 reports the heterogeneous effect of share of vocational students on language and math test scores of regular students, considering the educational level achieved by students at 9th grade. The specification includes covariates in each model. All specification contains school and time fixed effect. The standard errors are estimated clustering by school. Significance levels: 1% ***, 5% **, 10% *.

5.2 Effect on student's drop-out and retention

There are relatively few pieces of evidence of the long-run educational consequences of peers. Carrell, Hoekstra, and Kuka (2018), Anali and Peri (2017), and Bifulco et al (2014) are exceptions. We attempt to assess the long-run effect of reducing peer quality by estimating the impact on the rate of student dropout and the rate of student retention. Both variables have long-run consequences for secondary students.

We consider a school-level specification where the dependent variables are the rate of students' drop-out – i.e. the proportion of students that dropped out the school in each cohort – and the rate of students' retention – i.e. the ratio of students who remains in the same grade in each year. The specifications include school controls as the average class size and the proportion of teachers with tertiary. We also include school and time fixed effects.

Table 5 shows the results. Column 1 refers to the impact on school dropout considering all municipalities in Ceará, and column 2 presents the effect on school dropout just for small municipals. Column 3 refers to the impact on school retention for all municipals, and column 4 shows the effect on just small municipals. We restrict the sample for small municipalities in columns (2) and (4) because the stay at school can be affected by the local labor market opportunities.

The results suggest that reducing peer quality increase the likelihood of school drop-out in 5.3 percentual points, considering small municipalities (Column 2). On average, 11,4% of secondary students dropped out during secondary education in Ceará. Thus, the estimate represents an increase of almost 50% in the rate of school drop-out on average. The impact of

peer on student drop-out is large, suggesting that peer quality matters for longer-run outcomes. Nonetheless, in column 1, the effect of reducing peer quality is not significant, although it is positive.

In turn, peer quality also affects the rate of school retention significantly. As reported in columns 3 and 4, the impact of the share of vocational students on regular school retention is 4.511 percentage points considering all municipalities and 4.259 percentage points restricting to the small's one. On average, the rate of school retention is 7.08% during secondary education. This represents a rise of approximately 63% in the school retention.

Taken together, we conclude that reducing peer quality has relevant consequences for students' outcomes in both short and long-run. It is also interesting to note that the construction of vocational school has an unintended effect on students that were not directly affected by these schools. This highlights the importance of public policy design.

Table 5: Effect on the rate of school drop-out and retention

VARIABLES	(1)	(2)	(3)	(4)
	Dropout		Retention	
	All Municipalities	Small Municipalities	All Municipalities	Small Municipalities
Share of Vocational Students	3.978 (2.637)	5.350** (2.622)	4.511*** (1.486)	4.259*** (1.539)
Observations	2,930	2,000	3,893	2,674
R-squared	0.444	0.423	0.317	0.403
Municipal FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
School Controls	Y	Y	Y	Y

Notes: Table 6 reports the impact of share of vocational students on language and math test scores of regular students, considering the educational level achieved by students at 9th grade. The specification includes covariates in each model. All specification contains school and time fixed effect. The standard errors are estimated clustering by school. Significance levels: 1% ***, 5% **, 10% *.

6 ROBUSTNESS

We conduct several robustness tests to verify if the estimates are robust. First, to check if the estimates capture a spurious correlation among the share of students attending the vocational school and the regular students' outcomes, we realize a falsification test using placebo regression. In short, we verify if the share of vocational students in the municipality m , in the year t affects the students' outcomes of regular students in the same municipality, in the year $t - 1$. If the estimates are significant, then potentially the treatment effects are driven by short-run trends.

Table 6 reports the placebo test estimations. We show no significant effect on the share of vocational students on previous regular students' test scores. The exception is the column (1) for language test scores. This specification does not include any covariates. The inclusion of the control variables eliminates the significance, suggesting that the validity of the estimates is conditioning to covariates in such a case.

Next, we test if the results depend on the measure of vocational school's exposure. We do not expect that distinct measures of the vocational school's intensity yield contrasting estimates. We consider two alternative measures of vocational school's exposure. First, it is a binary variable indicating the year that some municipality received the first vocational school. In this case, the model is interpreted as a standard difference-in-difference estimation with dynamic treatment. Second, we define the proportion of vocational schools for each municipal-by-year. Table 7 shows the results for both alternatives measures. The estimates are not

insignificant and the signal of the parameters is negative, suggesting that the vocational school's exposure can negatively affect the students' outcomes.

Table 6: Falsification test

<u>Math</u>	(1)	(2)	(3)	(4)	(5)
Placebo Effect	-0.024 (0.016)	-0.007 (0.014)	0.029 (0.030)	0.038 (0.029)	0.041 (0.029)
<u>Language</u>	(1)	(2)	(3)	(4)	(5)
Placebo Effect	-0.033** (0.014)	-0.022 (0.031)	-0.022 (0.027)	-0.010 (0.031)	-0.008 (0.031)
No control	Y				
Regular control		Y	Y	Y	Y
Prior test scores			Y	Y	Y
Additional control				Y	Y
Non Linearities					Y

Notes: Table 6 reports falsification test. It tests if the share of vocational students in the municipality m , in the year $t - 1$ affects the students' outcomes of regular students in the same municipality, in the year $t - 1$. All specification contains school and time fixed effect. The standard errors are estimated clustering by school. Significance levels: 1% ***, 5% **, 10% *.

Finally, we test if the learning incentives depend on local economic opportunities. Big municipalities yield distinct local labor market incentives in comparison with small ones, potentially affecting students' effort. To test this possibility, we also estimate a model that ignores municipalities with more than 150 thousand inhabitants¹⁷.

Table 7: Alternative measures to vocational school exposure

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Math	Math	Math	Language	Language	Language
Binary Treatment	-0.0991*** (0.0142)	-0.0916*** (0.0143)	-0.0943*** (0.0146)	-0.0686*** (0.0122)	-0.0586*** (0.0119)	-0.0613*** (0.0121)
School Treatment	-0.000164** (0.000)	- (0.000)	- (0.000)	-0.000170*** (0.000)	- (0.000)	- (0.000)
School FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Student Controls	N	Y	Y	N	Y	Y
School Controls	N	N	Y	N	N	Y

Notes: Table 7 reports similar estimation of tables 3 and 4, however with alternative measures vocational school exposure. Binary treatment represents an indicator variable with one to municipalities m that receive a vocational school in time t , and School Treatment represent the share of vocational school in relation to regular ones. All specification contains school and time fixed effect. The standard errors are estimated clustering by school. Significance levels: 1% ***, 5% **, 10% *.

¹⁷ We perform additional robustness checks do not reported by concision. For more details, contact the authors by email.

7 MECHANISMS

Few papers address empirically the channels that explain the peer effects on education, exceptions are Duflo, Dupas, and Kremer (2011), and Lavy, Paserman, and Schlosser (2012). From a theoretical point of view, two concurrent theories attempt to predict the mechanisms of peer effect. First, reducing peer quality can affect classroom behavior, impacting student's effort to learning as in Lazear (2001). In turn, Duflo, Dupas, and Kremer (2011) argue that student composition can change teacher pedagogical practices, and thus affecting the quality of learning. Interestingly, theories predict contradictory effects for the importance of peers in secondary education.

This paper contributes to this discussion by assessing the channels of peer effects in an exploratory way. Using the contextual survey of SPAECE, we select 20 items referring to the following categories: student engagement, social skills, interest in learning, school well-being, noise at school, expectations, and time allocation. We test the effect of the proportion of students attending vocational school on these indicators¹⁸.

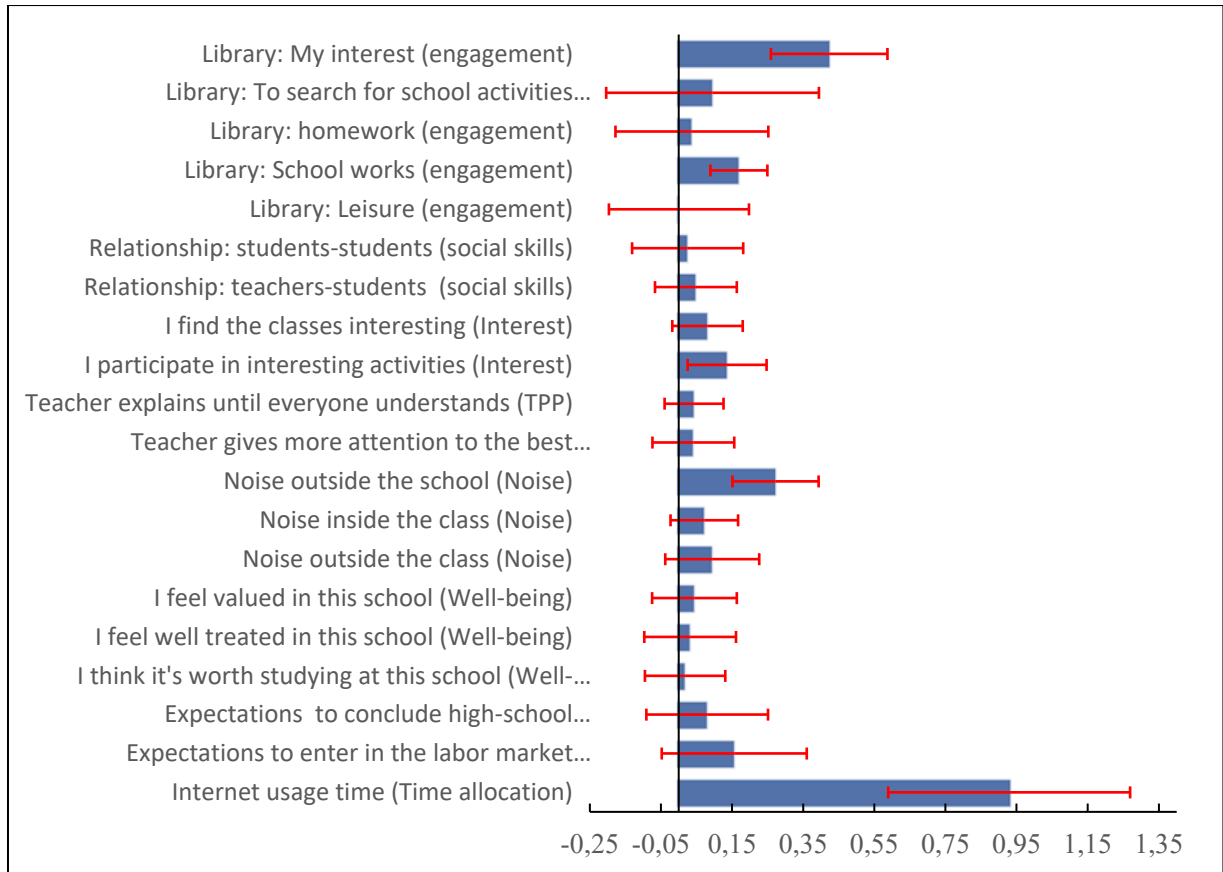
We consider a model that includes covariates as the age, gender, race, an indicator if the student's family attends a conditional cash-transfer program, and mother's education. Also, we include time and school fixed effects to control for time-varying and time-invariant unobservable factors. Figure 4 displays the estimates for each item. In parenthesis is the categories' names. For example, "*Library: My interest (engagement)*" refers to the question "*I go to the library to my own interest*" and is related to school engagement. All items are standardized, then the estimates represent standard deviations.

Only four variables present significant effects: going to the library by my interest, going to the library because of school works, the noise outside the school, and internet usage. The last two variables suggesting that peer quality affects student attention. The share of students attending vocational school affects the student time spending on the internet, i.e., reducing peer quality changes the student's time allocation. Another relevant factor is the noise outside the school, which is negatively affected by the share of vocational students. Both results suggest that reducing peer quality diverting the student focus on learning.

Interestingly, our results contrast with Duflo, Dupas, and Kremer (2011), and Lavy, Paserman, and Schlosser (2012) because we do not support the notion that a high proportion of low-achieving students induce teachers to modify their pedagogy and their personalized attention to better students. Students do not perceive that teacher diverting her attention to specific students.

¹⁸ Although the contextual survey of the SPAECE to be applied in all year of the sample, there are substantial difference among the items across the years. Therefore, we consider only the cohorts 2010-2013 and 2011-2014. To detailed information about this sample, see the Online Appendix available by email to the authors.

Figure 4: Effect of peer quality on potential mechanisms.



Notes: Figure 3 presents the impact of the share of vocational students on several variables that explains the mechanisms of why the reduction of peer quality affect students' outcomes. We regress each variable against additive school and fixed effects and the share of vocational students. The standard errors are estimated clustering by school level.

8 CONCLUSION

This paper investigates if peer composition affects secondary students' outcomes. To identify the causal effects, we exploit the variation of vocational schools in time and location across municipals of Ceará from 2008 to 2017. Vocational schools in Ceará realize an admission process to cherry-pick the high-achievers from primary education. This selective process reduces the share of high-achievers in regular secondary schools.

We find that regular students exposed to vocational school construction decrease their performance in math and language test scores. The effect is large, significant and presents small variations in different specification, suggesting that the estimates are reliable.

In addition, we also test the effect of peer composition on the rate of school drop-out and the rate of retention during secondary education using a school-level version of the empirical strategy. Our results indicate that reducing the peer quality increase the rate of students that drop-out the school (this result is restricted to large municipalities) and the rate of repeaters. We realize a battery of robustness' checks and we conclude that our estimates are unlikely to be biased.

Moreover, we are interested to understand the underlying mechanisms that explain the results. Using a contextual survey applied to students, we test different potential channels: student engagement, social skills, teachers' pedagogical practices, student's well-being, school and classroom noise, and student's time allocation.

The vocational school exposure affects student time allocation and outside school noise. Specifically, the share of students attending vocational schools increases the time spending by regular students on the internet. Therefore, we conclude that the mechanism that explains our results is the students' diversion, following the "bad apple" theory of Lazear (2001).

Finally, this paper contributes to literature investigating the unintended effect of some policies. In our case, the construction of selective public schools changes the composition of regular schools, reducing the quality of their peers. The effect harms the regular students affecting short and long-run outcomes. Policy-makers should account for this spillover effect when evaluating the overall impact of the vocational school expansion in Ceará.

REFERENCES

ANELLI, Massimo, and PERI, G. The Effects of High School Peers' Gender on College Major, College Performance and Income. *Economic Journal*, 2017.

ANGRIST, J. D. (2014). The Perils of Peer Effects. *Labour Economics*, 30, 98-108.

ATKIN, David, "Endogenous Skill Acquisition and Export Manufacturing in Mexico," *American Economic Review*, August 2016, 106 (8), 2046–85.

BIANCHI, N. The Indirect Effects of Educational Expansions: Evidence from a Large Enrollment Increase in University Majors, *Journal of Labor Economics*, v. 38 (3), 2020.

BIFULCO, Robert, FLETCHER, J.; OH, S. S. and ROSS, S. Do High School Peers Have Persistent Effects on College Attainment and Other Life Outcomes?" *Labour Economics* 29: p. 83–90, 2014.

BILLINGS S, DEMING DJ, ROSS S. Partners in Crime: Schools, Neighborhoods and the Formation of Criminal Networks, NBER Working Paper n° 21962, 2016.

BOSSAVIE, L. The Effect of Immigration on Natives' School Achievement: Does Length of Stay in the Host Country Matter? *Journal of Human Recourses*, 55(2), p. 733-766, 2020.

CARRELL, S.; HOEKSTRA, M. and KUKA, E. The long-run effects of disruptive peers. *The American Economic Review*, 108 (11), p. 3377-3415, 2018.

CARRELL, Scott E., and James E. WEST. Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors. *Journal of Political Economy* 118 (3): p. 409–32, 2010.

CARRILLO, B. Present Bias and Underinvestment in Education? Long-run effects of childhood exposure to booms in Colombia, *Journal of Labor Economics*, (forthcoming), 2020.

CHARLES, Kerwin Kofi, HURST, E., and NOTOWIDIGDO, M. Housing Booms and Busts, Labor Market Opportunities, and College Attendance, *American Economic Review*, October 108 (10), 2947–94, 2018.

DUCKWORTH, A. L., PETERSON, C., MATTHEWS, M. D., & KELLY, D. R. Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), p. 1087-1101, 2007.

- DUFLO, Esther. The medium run effects of educational expansion: Evidence from a large school construction program in Indonesia. *Journal of Development Economics*, 74(1), p. 163–197, 2004.
- DUFLO, E., DUPAS, P., and KREMER, M. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *The American Economic Review*, 101(5), 1739-1774, 2011.
- FALK, A. and ICHINO, A. Clean Evidence on Peer Effects. *Journal of Labor Economics*, 24 (1), 39-57, 2006.
- FORTIN, Bernard and YAZBECK, Myra, Peer effects, fast food consumption and adolescent weight gains. *Journal of Health Economics*, vol. 42(C), p. 125-138, 2015
- GEORGANAS, S.; TONIN, M and VLASSOPOULOS, M. Peer pressure and productivity: the role of observing and being observed. *Journal of Economic Behavior and Organization*. Vol 11(c), p. 223-232, 2015.
- GILRAINE, M; MACRTENEY, H. and McMILLAN, R. Education Reform in General Equilibrium: Evidence from California's Class Size Reduction, NBER Working Paper No. 24191, 2018.
- GLEWWE, Paul and KASSOUF, Ana Lucia, The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil, *Journal of Development Economics*, vol. 97(2), p. 505-517, 2012.
- GOULD, Eric D., LAVY, V. and PASERMAN, M. Daniele. Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence. *Economic Journal* 119 (540): p.1243–69, 2009.
- HOXBY, C. M., and WEINGARTH, G. Taking Race out of the Equation: School Reassignment and the Structure of Peer Effects: Working paper, 2005.
- IMBERMAN, Scott A., KUGLER, A. and SACERDOTE, B. Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees. *American Economic Review* 102 (5): p. 2048–82, 2012.
- JACKSON, K. Teacher Quality at the High-School Level: The Importance of Accounting for Tracks, *Journal of Labor Economics*, V. 32(4), 2014.
- LAVY, V., PASERMAN, M. D., and SCHLOSSER, A. Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. *The Economic Journal*, 122(559), p. 208-237, 2012.
- LAZEAR, E. P. Educational Production. *The Quarterly Journal of Economics*, 116(3), p. 777-803, 2001.
- MANSKI, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60 (3), 531.
- SACERDOTE, B. Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward? *Annual Review of Economics*, 6 (1), 253-272, 2014.

SHILTZ, F.; MAZREKAJ, D.; HORN, D. and De WITTE, K. Does it matter when your smartest peers leave your class? Evidence from Hungary, Labour Economics, Volume 59, p. 79-91, 2019.

APPENDIX

Table A1: **Descriptive statistics**

Variables	Treated		Untreated		Difference
	Average	SD	Average	SD	
<i>Test Scores</i>					
Test Scores Math 9 ^o	236.80	44.50	236.70	44.83	0,10
Test Scores Portuguese 9 ^o	229.00	45.75	227.72	45.50	1,28
Test Scores Math 3 ^o	259.24	49.08	258.72	49.82	0,52
Test Scores Portuguese 3 ^o	258.07	47.08	251.48	46.79	6,59*
<i>Race</i>					
Black	0.223	0.416	0.130	0.336	0,09*
Brown	0.125	0.331	0.081	0.274	0,04*
White	0.206	0.404	0.134	0.341	0,07*
Other	0.443	0.496	0.653	0.475	0,21*
<i>Mother education</i>					
Incomplete Primary	0.023	0.150	0.012	0.109	0,01*
Primary	0.169	0.289	0.093	0.216	0,07*
Secondary	0.055	0.228	0.031	0.174	0,11*
Higher	0.019	0.139	0.011	0.106	0,008
<i>Others</i>					
Conditional Cash-Transfer	0.754	0.430	0.742	0.437	0,01
Public Transport	0.290	0.453	0.510	0.499	-0,22
Girl	0.551	0.497	0.566	0.495	0,01

ANÁLISE DA INCIDÊNCIA E DO PERÍODO DE AFASTAMENTO POR ACIDENTES DE TRABALHO OU DOENÇAS OCUPACIONAIS NO ESTADO DO CEARÁ

Mateus Mota dos Santos*

Isabela Braga Sales**

David Herbster Ferraz***

RESUMO

O Ceará é o terceiro estado de maior incidência de acidentes de trabalho na região Nordeste do Brasil. Somente em 2017, quase 11 mil trabalhadores foram vítimas de acidentes de trabalho no estado. Diante disso, esta pesquisa busca analisar como as características individuais e ocupacionais do profissional e as características da empresa se relacionam com a probabilidade de afastamento da atividade laboral devido à ocorrência de acidentes ou doenças ocupacionais no estado do Ceará. Busca-se também averiguar como esses fatores afetam a duração do período de afastamento do trabalho causado por esses motivos. Para tanto, foram utilizadas informações extraídas da Relação Anual de Informações Sociais (RAIS), referente ao ano de 2017, além de um modelo de regressão logística e de técnicas de análise de dados de sobrevivência. Os resultados revelam que as chances de afastamento por acidente ou doenças ocupacionais são maiores para os homens, mas que o período de afastamento tende a ser mais elevado para as mulheres, e que os trabalhadores mais velhos não são apenas mais propensos a se envolverem em acidentes, mas também necessitam de um período de tempo maior para a recuperação. Além disso, verificou-se que os trabalhadores com ensino superior são menos propensos a acidentes de trabalho. No entanto, os graduados que atuam em empregos que exigem abaixo do ensino superior são duas vezes mais propensos ao afastamento do que seus pares adequadamente alocados, ou seja, engajados em empregos condizentes com o seu nível de escolaridade.

Palavras-chave: Acidentes de Trabalho; Doenças Ocupacionais; Estado do Ceará.

ABSTRACT

Ceará is the third state with the highest incidence of occupational accidents in the Northeast region of Brazil. In 2017 alone, almost 11,000 workers were victims of occupational accidents in the state. Therefore, this research seeks to analyze how the individual and occupational characteristics of the professional and the characteristics of the company are related to the probability of leaving work due to the occurrence of accidents or occupational diseases in the state of Ceará. It also seeks to ascertain how these factors affect the length of the period of absence from work caused by these reasons. For this purpose, information extracted from the Annual Report on Social Information (RAIS) for the year 2017 was used, in addition to a logistic regression model and techniques for analyzing survival data. The results reveal that the chances of being on leave due to accidents or occupational illnesses are higher for men, but that the period of leave tends to be higher for women, and that older workers are not only more likely to be involved in accidents, but they also need a longer recovery time. In addition, it was found that workers with higher education are less prone to accidents at work. However, graduates who work in jobs requiring less than higher education are twice as likely to be dismissed as their appropriately allocated peers, that is, engaged in jobs consistent with their level of education.

Keywords: Accidents at Work; Occupational Diseases; State of Ceará.

Classificação JEL: J28, J81

* Doutorando em Economia – CAEN/UFC. E-mail: matheus12_mota@hotmail.com. Tel: (88) 992974415.

** Doutoranda em Economia – CAEN/UFC. E-mail: isabela_braga_sales@hotmail.com.

*** Mestre em Economia – CAEN/UFC. E-mail: daviferraz@caen.ufc.br.

1 INTRODUÇÃO

A ocorrência de acidentes de trabalho constitui um importante problema de saúde pública, pois além de suas consequências produtivas, financeiras, previdenciárias e jurídicas, afeta negativamente a saúde física e psíquica dos trabalhadores acidentados, seus aspectos sociais, profissionais e familiares (GUIMARÃES; GRUBITS, 2004).

De acordo com o artigo 19 da Lei nº 8.213, de 24 de julho de 1991, entende-se por acidente de trabalho o evento decorrente do exercício laboral a serviço da empresa, que acarreta na morte, perda ou redução temporária ou permanente da capacidade de trabalho do profissional acidentado. Os sinistros podem ser classificados em acidentes típicos, decorrentes do exercício da própria atividade laborativa, acidentes de trajeto, ocorridos a caminho da empresa, de casa para o trabalho e vice-versa, e doenças ocupacionais, associadas ao ofício e às condições de trabalho do indivíduo.

Segundo dados do Anuário Estatístico de Acidentes do Trabalho (2017), somente em 2017, foram registrados no Brasil 549.405 acidentes do trabalho dos quais 450.614 foram informados na Comunicação de Acidentes de Trabalho (CAT). Comparado com 2016, o número de acidentes de trabalho apresentou uma redução de 6,2%. Do total de acidentes com CAT registrada, 75,5% foram acidentes típicos; os de trajeto representaram 22,3%; e as doenças ocupacionais 2,2%. No Brasil, o Ceará é o 12º estado de maior incidência de acidentes de trabalho, com aproximadamente 11 mil casos, o que corresponde a uma média de 30 acidentes por dia. Na região Nordeste, o estado ocupa a terceira posição no ranking de acidentes de trabalho, atrás apenas da Bahia, com 16.332 acidentes, e Pernambuco, com 14.417 casos.

Esses eventos podem demandar assistência médica ou afastamento (menos ou mais de 15 dias), causar incapacidade permanente ou ainda levar o profissional a óbito e, consequentemente, implicam em custos. Esses, por sua vez, são compostos por gastos com saúde, seguros e previdência. De acordo com a International Labour Organization (2003), estima-se que os acidentes de trabalho podem causar perdas de magnitude de 4% e 10% do PIB para países desenvolvidos e em desenvolvimento, respectivamente. No Brasil, apenas os gastos previdenciários que incluem auxílio-doença, aposentadoria por invalidez, pensão por morte e auxílio-acidente somaram R\$ 13,2 bilhões no ano de 2017. Ademais, foram contabilizados mais de 42,5 milhões de dias de trabalho perdidos em decorrência de afastamentos. Para o Ceará, durante o mesmo período, os gastos previdenciários somaram aproximadamente R\$ 200 milhões, e o tempo de trabalho perdido em decorrência do acidente de trabalho sofrido foi em torno de 769,7 mil dias (SMARTLAB, 2020).

Acidentes de trabalho podem ser evitados e é de suma importância a prática de estratégias que promovam segurança no ambiente laboral (BENA et al., 2009; ADERAW et al., 2011). Os custos já mencionados, além do desgaste emocional das pessoas próximas ao acidentado, incorrem em enorme prejuízo para a sociedade como um todo. Dessa forma, o estudo do tema é relevante não apenas para as empresas, mas também para os formuladores de políticas cujo objetivo reside na prevenção de tais infortúnios. E isso pode ser feito por meio do reconhecimento de grupos de trabalhadores que se encontram mais propensos a lesões decorrentes do exercício de atividade de trabalho.

Diante disso, este estudo tem por objetivo analisar como as características individuais e ocupacionais do profissional e as características da empresa se relacionam com a probabilidade de afastamento da atividade laboral devido à ocorrência de doenças ocupacionais ou de acidentes ocorridos no ambiente de trabalho no estado do Ceará. Busca-se também averiguar como esses fatores afetam a duração do período de afastamento do trabalho causado por esses mesmos motivos. Para isso, além do método de regressão logística, utilizou-se de

técnicas de análise de sobrevivência aplicadas aos dados da Relação Anual de Informações Sociais (RAIS), referentes ao ano de 2017.

A estrutura do trabalho foi dividida em cinco seções, incluindo essa. As demais são: revisão de literatura, onde são expostos diversos aspectos acerca de acidentes de trabalho e sua relação com os mesmos; metodologia, na qual são apresentados os dados, variáveis e métodos utilizados; resultados, cujo conteúdo aborda os achados provenientes das investigações realizadas; e, por fim, as considerações finais.

2 REVISÃO DE LITERATURA

Por ser um tema relevante e caracterizado como um problema de saúde pública (AHN; BENA; BAILER, 2004), diversos estudos, tanto a nível nacional quanto internacional, buscaram compreender as causas de acidente de trabalho bem como maneiras de mitigá-lo. As más condições laborais (GARCÍA-HERRERO et al. 2012; CIONI; SAVIOLI, 2016; ROMMEL et al., 2016), a falta de treinamento em saúde e segurança ocupacional (BENA et al., 2009; ADERAW et al., 2011), longas horas de trabalho e menos experiência (TADESSE; KUMIE, 2007; DONG et al., 2015), baixo nível de educação (PIHA et al., 2013; CAVALCANTE, 2013; JAFARI et al., 2019), e ser do sexo masculino (BHATTACHERJEE et al., 2003; NAKATA et al., 2006) são frequentemente listados como os principais fatores de risco de lesões ocupacionais.

A influência das condições de trabalho sobre a prevalência de acidentes foi investigada por García-Herrero et al. (2012), em um estudo para a Espanha. A realização da pesquisa contou com o uso de informações extraídas do *VI National Survey of Working Conditions* de 2007, e de técnicas probabilísticas multivariadas. Modelos de redes bayesianas foram desenvolvidos com o objetivo de medir a influência exercida pelas condições de trabalho, agrupadas em condições higiênicas, condições ergonômicas, e demandas de trabalho, sobre sintomas físicos e psicológicos relacionados ao trabalho e sobre as ocorrências de acidentes ocupacionais. Os resultados do modelo revelam que as más condições higiênicas duplicam a probabilidade de acidentes de trabalho, aumentando-a de 10,46% para 21,53%, e que a manutenção de boas condições higiênicas ou de melhorias nas condições ergonômicas reduziria as taxas de acidentes em 5%. Também evidenciam que situações ergonômicas ruins aumentam os sintomas físicos de 32,32% para 47,66%, e as altas demandas de trabalho, como atenção requerida, esforço necessário para realização da atividade laboral, elevam os sintomas psicológicos de 10,49% para 18,43%.

Rommel et al. (2016) estudaram a influência de um conjunto de características sociodemográficas, relacionadas ao emprego e à saúde, sobre a ocorrência de acidentes de trabalho, utilizando informações extraídas do *German Health Update 2010*, uma pesquisa por telefone representativa da população adulta da Alemanha. Estimativas obtidas a partir da análise de regressão logística revelaram que, as chances de envolvimento em acidentes foram 78% maior em atividades que exigiam o carregamento de pesos e que, trabalhar em posturas inadequadas esteve associado a um aumento de 46% nos riscos de acidentes. Também evidenciam que o estresse e o trabalho sob pressão aumentam a probabilidade de acidentes em 48% e 41%, respectivamente.

A realização de treinamento em saúde e segurança no trabalho também desempenha um papel importante na redução das taxas de acidentes. A participação dos trabalhadores em programas de treinamento poderá aumentar a conscientização destes sobre os diferentes tipos de riscos aos quais estão sujeitos, além de melhorar a compreensão dos requisitos de segurança no trabalho e aprimorar suas competências e habilidades profissionais associadas a tarefas específicas, e com isso reduzir os riscos de acidentes. Bena et al. (2009), por exemplo, avaliaram

o impacto de um programa de treinamento em saúde e segurança sobre a ocorrência de lesões em trabalhadores manuais engajados em atividades nos canteiros de obras da linha ferroviária de alta velocidade Torino-Novara, na Itália. O objetivo do programa era o de conscientizar os trabalhadores sobre os riscos de acidentes específicos no local de trabalho, fornecendo-lhes competência adequada para prevenção e controle de tais riscos. Sua realização se deu em diferentes etapas, a partir de um módulo básico e outros direcionados a funções de trabalho mais específicas. Os resultados obtidos com a realização de análises pré e pós-treinamento mostram que a incidência de lesões caiu 16% após a conclusão do módulo de treinamento básico e, após os módulos mais específicos, diminuiu em 25%. Ademais, a análise de modelo de série temporal mostrou que o treinamento resultou em uma redução de 6% nas taxas de lesões nos canteiros de obras.

A forma como as características profissionais dos trabalhadores se relacionam com os riscos de acidentes no ambiente de trabalho também tem sido investigada pela literatura. Na Etiópia, Tadesse e Kumie (2007) avaliaram a prevalência de lesões em trabalhadores empregados em pequenas e médias indústrias. Estes observaram que as chances de acidentes em trabalhadores com cinco ou menos anos de experiência na mesma empresa, são 1,5 vezes maiores do que as de trabalhadores com mais tempo de serviço. Os que trabalhavam por mais de 48 horas semanais, e que não eram supervisionados regularmente, também assumiram maior probabilidade de se envolverem em acidentes, 32% e 39% a mais que seus pares que trabalhavam por menos horas e que eram supervisionados com frequência, respectivamente.

Nos Estados Unidos, Dong et al. (2015) também encontram resultados semelhantes. Usando informações do *National Longitudinal Survey of Youth* de 1988-2000, e com o auxílio de modelos de regressão logística multivariada, os autores evidenciam que a existência de vários empregos e a exposição a longas horas de trabalho, aumenta as chances de lesões ocupacionais. Trabalhadores empregados em cinco ou mais empregos tinham duas vezes mais probabilidade de se envolverem em acidentes do que aqueles com um ou dois empregos, e o desempenho de atividades laborais por mais de dez horas diárias aumenta as chances de acidentes em 15% em relação ao exercício de oito horas diárias. Tal resultado pode estar associado ao maior desgaste físico e mental, maior fadiga e insatisfação daqueles submetidos a maiores jornadas de trabalho.

Quanto às características relacionadas à empresa, Holte et al. (2015) observaram que a prevalência de acidentes de trabalho varia sistematicamente entre estabelecimentos de tamanhos distintos. Estes estudaram as diferenças nos riscos de lesões entre aprendizes em empresas de diferentes tamanhos e em diferentes ramos de construção, no oeste da Noruega. Suas estimativas de risco relativo identificaram que os riscos de acidentes são significativamente maiores em empresas de menor porte. A prevalência de lesões em aprendizes foi maior em empresas que empregavam de 10 a 19 funcionários, em comparação com aquelas de maior porte, acima de 50 funcionários, com razão de risco relativo estimada em 1,56. Ao estratificar a análise para o comércio de construção e de eletricidade, os autores encontraram que, no primeiro e segundo grupo, respectivamente, os aprendizes em empresas que empregavam 20-49 e 10-19 trabalhadores tiveram uma razão de prevalência duas vezes maior para se ferir durante o treinamento do que os aprendizes em empresas com 50 ou mais funcionários.

Em um estudo para a Espanha, López-Arquillos e Rubio-Romero (2016) também evidenciam que o tamanho do estabelecimento, medido pela quantidade de funcionários, influencia a ocorrência de acidentes de trabalho. Seus resultados confirmam que os riscos de lesões como queimaduras, feridas superficiais, fraturas ósseas, escaldamento e congelamento, são maiores em empresas com cinco ou menos funcionários, sendo mais fortes para as causas de queimaduras, escaldamento e congelamento, onde as chances de incidência são 37,6% maiores em relação às empresas de maior porte. Segundo os autores, uma possível razão para

esses resultados está na dificuldade de controle dos riscos de lesões e de acidentes enfrentada pelas empresas menores, em comparação com as maiores, em virtude da limitação de recursos materiais e humanos. Além disso, as empresas de maior porte normalmente contam com uma gestão das questões de prevenção mais sólida, e dispõem de melhores atividades de treinamento e de informações se comparadas às empresas de menor capacidade, como já destacado por Fabiano et al. (2004), em um estudo para a Itália. Os autores também constataram que o gênero pode ser considerado um fator de risco ou de proteção, a depender do tipo de lesão considerada. As mulheres, por exemplo, são menos propensas a sofrerem queimaduras e feridas superficiais, e mais propensas a sofrerem luxações ou entorses, em relação aos homens.

Inúmeras pesquisas evidenciam que as chances de acidentes de trabalho são maiores para trabalhadores mais jovens e do sexo masculino (BHATTACHERJEE et al., 2003; NAKATA et al., 2006; ADERAW et al., 2011). No nordeste da França, por exemplo, Bhattacharjee et al. (2003) encontram que ser do sexo masculino e ter menos de 29 anos de idade aumenta as chances de lesões ocupacionais em 1,99 e 1,7 vezes, respectivamente. Na cidade de Yashio, Saitama, Japão, Nakata et al. (2006) mostram que os homens são 3,46 vezes mais propensos a sofrerem lesões relacionadas ao trabalho do que as mulheres, enquanto o grupo de trabalhadores mais jovens, com idade entre 16 e 29 anos, são duas vezes mais propensos, se comparado aos mais velhos, acima de 59 anos. Ao contrário desses estudos, um corpo crescente de pesquisa aponta para uma relação positiva entre idade e acidente de trabalho (BANDE; LÓPEZ-MOURELO, 2015; HASEBE; SAKAI, 2018), chamando atenção para as implicações do envelhecimento da força de trabalho na saúde e segurança ocupacional dos trabalhadores. Como destacam Hasebe e Sakai (2018), apesar do acúmulo de experiência e habilidade obtida ao longo da carreira, o declínio da função física ou cognitiva em razão do envelhecimento pode contribuir para o aumento dos riscos de lesões ou doenças graves relacionadas ao trabalho.

Diferentes estudos também destacam com bastante frequência que as pessoas com status educacional e ocupacional mais elevado se envolvem menos em acidentes de trabalho do que aquelas menos escolarizadas e engajadas em empregos que exigem menos qualificação (OH; SHIN, 2003; PIHA et al., 2013; JAFARI et al., 2019; GEBREMESKEL; YIMER, 2019). Piha et al. (2013) analisaram as diferenças socioeconômicas na ausência por acidentes de trabalho na cidade de Helsinque, capital da Finlândia. Educação, classe ocupacional e renda dos trabalhadores foram usadas pelos autores como medidas de posição socioeconômica. Os resultados obtidos com a análise de regressão de Poisson confirmam a existência de diferenças socioeconômicas nos riscos de afastamento por lesões de trabalho, sendo mais elevadas entre as classes ocupacionais do que entre as medidas de educação e renda individual. As estimativas para as mulheres mostram que os riscos de acidentes são duas vezes maiores para atividades manuais do que para gestores e profissionais. Entre os homens, esse diferencial é ainda maior, onde os trabalhadores manuais são quase cinco vezes mais propensos ao afastamento por acidentes de trabalho. As diferenças de escolaridade também foram significativas, e apontam que trabalhadores com baixo nível educacional são mais propensos a se envolverem em acidentes do que os mais instruídos. Mulheres e homens com baixa escolaridade (escolaridade obrigatória) são 1,39 e 1,81 vezes mais propensos a afastamento por causa de acidentes se comparados aos seus pares de alta educação (ensino médio completo). Jafari et al. (2019), a partir de uma amostra de trabalhadores iranianos selecionados aleatoriamente da *Arya Sasol Petrochemical Company*, na cidade de Bushehr, durante o ano de 2016, observaram que, mesmo após o controle de um conjunto de características pessoais e relacionadas ao trabalho, a escolaridade do indivíduo influencia de forma significativa as chances de acidentes de trabalho. Suas estimativas apontam que os trabalhadores com formação acadêmica são aproximadamente 83% menos propensos a sofrerem lesões ocupacionais em relação aos não acadêmicos.

No Brasil, a problemática do acidente de trabalho tem sido investigada por muitos pesquisadores (CAVALCANTE, 2013; RIOS et al., 2015; SOARES et al., 2018; GONÇALVES et al., 2018), os quais buscaram identificar os potenciais fatores de riscos associados à ocorrência de lesões e de doenças provenientes do trabalho. Seus resultados corroboram os achados da maioria das pesquisas internacionais discutidas anteriormente.

Visando analisar de que maneira as características pessoais e ocupacionais dos trabalhadores formais estão relacionadas com a probabilidade de acidente de trabalho na Região Nordeste do Brasil, Cavalcante (2013), utilizando dados da RAIS de 2011 e modelo de regressão logística, mostrou que as chances de acidentes ocupacionais diminuem com a idade e com maiores níveis de escolaridade. Também observou que os homens são aproximadamente duas vezes mais propensos a se envolverem em acidentes de trabalho do que as mulheres, e que as pessoas que atuam no setor de construção civil são 1,33 vezes mais propensas ao afastamento por acidente de trabalho do que aquelas engajadas em atividades pertencentes aos demais setores.

Rios *et al.* (2015) investigaram a incidência de acidentes de trabalho não fatais em trabalhadores informais do comércio, no município de Jequié, Bahia, no ano de 2013. Os resultados da análise de regressão logística revelaram que os riscos de acidentes são maiores para trabalhadores do sexo masculino, jovens, comerciantes de carnes, e para aqueles que realizavam esforço físico mais elevado. Os autores estimam que as chances de acidente para os homens superam as das mulheres em 61%, e que os trabalhadores jovens são quase cinco vezes mais propensos a se envolverem em acidentes no ambiente de trabalho.

Em uma pesquisa mais recente, Soares et al. (2018) analisaram a influência do estresse físico e psicossocial, e de fatores relacionados à saúde, na ocorrência de acidentes ocupacionais. Com base em informações da Pesquisa Nacional de Saúde de 2013, e da análise de regressão logística, os autores mostram que os trabalhadores ocupados em três ou mais empregos são aproximadamente três vezes mais propensos a se envolverem em acidentes de trabalho, em comparação com aqueles engajados em um único emprego. Os trabalhadores com excesso de horas trabalhadas podem estar sujeitos à exaustão, fadiga, e estresse, que, por sua vez, aumentam os riscos de acidentes. Indicadores relacionados ao trabalho, como a exposição a material biológico e a radioatividade também influenciam esses resultados, aumentando as chances de acidentes em aproximadamente duas e quatro vezes, respectivamente. Por fim, suas estimativas comprovam que os sinistros também estão associados ao comportamento individual do acidentado. As chances de envolvimento em acidentes são aproximadamente cinco e três vezes maiores para os indivíduos que consomem bebidas alcoólicas e que se encontram acima do peso, respectivamente, corroborando os achados de Rommel et al. (2016). O consumo de bebidas alcoólicas pode reduzir a capacidade produtiva do indivíduo e a segurança no espaço de trabalho, através da redução do equilíbrio e reflexos do indivíduo, podendo até mesmo pôr em risco à segurança dos demais.

Ainda para o Brasil, Gonçalves et al. (2018) estudaram os riscos ocupacionais e o perfil dos trabalhadores associado à ocorrência de acidentes de trabalho em uma empresa da indústria metalomecânica da macrorregião sul catarinense, ao longo do período 2007-2015. As taxas de acidentes foram mais elevadas em períodos de recessão econômica, 2008 e 2012, e nos grupos de trabalhadores do sexo masculino, com idade entre 18 e 29 anos, casado ou em uma união estável, com ensino superior incompleto e com menos de cinco anos de serviço. Também encontram que os riscos de afastamento por mais de quinze dias, em decorrência de acidentes no ambiente de trabalho, são 30% maiores para os funcionários que trabalhavam na empresa há menos de um ano, destacando a importância da experiência para a prevenção de acidentes. Por fim, os autores destacam que as lesões mais frequentes ocorridas no local de trabalho foram distensões, contusão, escoriações, corte e amputações, o que chama a atenção para a necessidade de uma fiscalização e medidas de segurança mais eficazes no ambiente de trabalho.

3 METODOLOGIA

3.1 Base de Dados e Descrição das Variáveis

Para a realização da presente pesquisa foram utilizados dados extraídos da RAIS de 2017, divulgados pelo Ministério do Trabalho e Emprego. A base contempla informações anuais acerca das características individuais e ocupacionais dos trabalhadores formalmente ocupados, além da empresa para a qual prestam serviços, construídas a partir das declarações dos próprios empregadores. Serão considerados na análise apenas trabalhadores com idade entre 18 e 70 anos que foram afastados ou não de sua atividade laboral no estado do Ceará em 2017. Uma síntese descritiva das variáveis consideradas na análise empírica é apresentada no Quadro 1.

Quadro 1 – Descrição das variáveis

Variável	Descrição
Variáveis Dependentes	
AcidenteTrabalho	1 = se foi afastado do emprego por motivos de acidentes ocorridos no local de trabalho ou doenças ocupacionais; 0 = caso não tenha sido afastado do trabalho ou que o foi por outros motivos
DiasAfastamentoTrabalho	Quantidade de dias de afastamento em decorrência de acidentes ocorridos no local de trabalho ou de doenças ocupacionais
Variáveis Explicativas	
Características Individuais	
Sexo	1 = Homem; 0 = Mulher
Cor	1 = Branco; 0 = Não-branco
Idade	Idade do indivíduo
FundIncompleto*	1 = se possui escolaridade aquém do ensino fundamental completo; 0 = c.c.
FundCompleto	1 = se possui ensino fundamental completo; 0 = c.c.
MédioCompleto	1 = se possui ensino médio completo; 0 = c.c.
SuperiorCompleto	1 = se possui ensino superior completo; 0 = c.c.
PortDeficiência	1 = se possui alguma deficiência; 0 = se não possui deficiência
Características Ocupacionais	
Experiência	Quantidade de anos trabalhando no mesmo emprego
PrimeiroEmprego	1 = se for o primeiro emprego; 0 = c.c.
FormMinCBO	1 = se a ocupação exige ensino superior; 0 = se a ocupação exige abaixo do ensino superior
HrsTrabalho	1 = se trabalha 40 horas semanais ou menos; 0 = se trabalha acima de 40 horas semanais
Renda00	1 = até um salário mínimo; 0 = c.c.
Renda01	1 = entre 1,01 e três salários mínimos; 0 = c.c.
Renda02	1 = entre 3,01 e cinco salários mínimos; 0 = c.c.
Renda03*	1 = acima de cinco salários mínimos; 0 = c.c.
Características da Empresa	
MicroEmp	1 = Microempresa; 0 = Demais tamanhos
EmpresaPP	1 = Empresa de Pequeno Porte; 0 = Demais tamanhos
EmpresaMP	1 = Empresa de Médio Porte; 0 = Demais tamanhos
EmpresaGP*	1 = Empresa de Grande Porte; 0 = Demais tamanhos
Setores Econômicos	
Agricultura Pesca	1 = Setores agropecuário e de pesca; 0 = Demais setores
Indústria	1 = Setor industrial; 0 = Demais setores
ConstCivil*	1 = Setor de construção civil; 0 = Demais setores
Comércio	1 = Setor de comércio; 0 = Demais setores
Serviços	1 = Setor de serviços; 0 = Demais setores

Fonte: Elaboração própria.

Nota: * variável utilizada como categoria de referência.

Dois exercícios serão realizados neste estudo. No primeiro, busca-se entender como as características individuais e ocupacionais do trabalhador e da empresa se relacionam com a probabilidade de afastamento por acidente de trabalho ou doença ocupacional no estado do Ceará. Nesse caso, a variável dependente é binária e assume valor unitário caso o indivíduo tenha sofrido algum afastamento do emprego causado por acidentes ocorridos no ambiente de trabalho ou doenças ocupacionais no ano de 2017 e, zero caso contrário. O segundo exercício concentra-se na análise da gravidade dos acidentes no estado, medida aqui pelo período de afastamento do trabalho em decorrência de acidentes ou de doenças relacionadas ao emprego. Para isso, usa-se a quantidade de dias de trabalho perdidos devido a essas causas como variável dependente.

A escolaridade mínima requerida pela ocupação, representada aqui pela variável FormMinCBO, também será utilizada como controle, pois a prevalência de acidentes pode diferir de forma significativa entre ocupações que exigem altos ou baixos níveis de qualificação para o desempenho de suas atividades. O nível de escolaridade formal exigido pelo emprego foi obtido a partir do código da Classificação Brasileira de Ocupações (CBO) de 2002, variável presente no banco de dados da RAIS. A CBO 2002 agrega os empregos segundo a similaridade de suas funções e habilidades cognitivas necessárias à realização de suas atividades. Através desta, foi possível identificar as ocupações que exigem como escolaridade mínima o ensino superior e aquelas que demandam escolaridade aquém desse nível.

Tendo em vista que os riscos de acidente diferem sistematicamente entre estabelecimentos de tamanhos distintos (HOLTE et al., 2015; LÓPEZ-ARQUILLOS; RUBIO-ROMERO, 2016), as empresas cearenses serão classificadas em micro, pequena, média e de grande porte, a partir de informações do Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE), conforme reportado no Quadro 2.

Quadro 2 – **Porte da empresa**

Porte	Comércio e Serviços	Indústria
Microempresa	Até 9 funcionários	Até 19 funcionários
Empresa de Pequeno Porte	10 a 49 funcionários	20 a 99 funcionários
Empresa de Médio Porte	50 a 99 funcionários	100 a 499 funcionários
Empresa de Grande Porte	No mínimo 100 funcionários	No mínimo 500 funcionários

Fonte: Elaborado a partir de informações do SEBRAE.

3.2 Estratégias Econométricas

3.2.1 Regressão Logística

Visando identificar os principais fatores de risco e de proteção associados à probabilidade de afastamento do trabalho por doença ou acidentes de trabalho no estado do Ceará, a análise de regressão logística será utilizada como estratégia empírica. Assim, seguindo Greene (2012), a probabilidade de um trabalhador ser afastado pode ser escrita por:

$$Prob(Y = 1|X) = 1 - Prob(Y = 0|X) = \frac{\exp(X'\beta)}{1 + \exp(X'\beta)} = \Lambda(X'\beta) \quad (1)$$

onde Y é a variável dependente de interesse, a qual assume valor um caso o indivíduo tenha sido afastado do trabalho devido a acidentes ou doenças ocupacionais, e zero caso contrário; o vetor X corresponde ao conjunto de características individuais e profissionais do indivíduo, além de características relacionadas a empresa; e a notação $\Lambda(X'\beta)$, indica a função de distribuição logística cumulativa.

Uma vez que os coeficientes estimados no modelo não representam diretamente a magnitude da mudança na probabilidade de acidente no trabalho resultante de alterações

unitárias em uma variável independente de interesse, enquanto as demais são mantidas constantes, os efeitos marginais dos regressores serão calculados:

$$\frac{\partial E[Y|X]}{\partial X} = \Lambda(X'\beta)[1 - \Lambda(X'\beta)]\beta \quad (2)$$

Outra forma de dar significado econômico aos resultados do modelo logit é estimá-lo em termos de razões de chances, $\frac{Prob(Y = 1|X)}{Prob(Y = 0|X)} = \exp(X'\beta)$, comparando a probabilidade de sucesso com a de fracasso, considerando a seguinte função de regressão:

$$g(X) = \ln\left(\frac{Prob(Y = 1|X)}{Prob(Y = 0|X)}\right) = X'\beta \quad (3)$$

Caso o valor do coeficiente estimado seja maior que a unidade, após uma variação unitária em uma variável explicativa de interesse, a probabilidade de afastamento do trabalho devido a acidentes ou doenças é maior do que a probabilidade de não afastamento. No caso em que a variável independente é dicotômica, o resultado da estimativa é interpretado em termos da mudança de uma categoria para a outra.

3.2.1 Análise de Sobrevivência

Esta pesquisa também busca analisar o efeito das características individuais e profissionais do indivíduo, e da empresa, na duração do período de afastamento do trabalho em decorrência de acidentes ou doenças ocupacionais no estado do Ceará. Para isso, usa-se o método de Análise de Sobrevivência, o qual consiste em um conjunto de processos estatísticos de análise de dados nos quais a variável de interesse é o tempo transcorrido até que um estado inicial específico seja modificado, e essa alteração corresponde ao evento de interesse. Neste estudo, o afastamento do trabalho por lesão ou doença representa o estado inicial, e o fim do período de afastamento, o evento de interesse. Formalizando a ideia, seja T uma variável aleatória que representa o tempo de afastamento do trabalho, medida em dias, e t um período específico de T . A probabilidade de se observar um período de afastamento do trabalho igual ou maior que t , é dado pela função de sobrevivência $S(t) = 1 - F(t) = \Pr(T \geq t)$, em que, a função de distribuição cumulativa, representada por $F(t)$, mede a probabilidade de ausência do trabalho por um período menor que t .

Um problema bastante comum na análise de sobrevivência é a presença de censura nos dados. No caso desta pesquisa, ocorre censura à direita devido à limitação do período de análise (12 meses), o qual não permite conhecer com exatidão a duração do período de afastamento para aqueles que não retornaram ao trabalho antes de um ano, após serem afastados por motivos de acidentes ou doenças ocupacionais¹. Na presença de dados censurados, geralmente usa-se o método de Kaplan-Meier para estimar a função de sobrevivência (KAPLAN; MEIER, 1958), o qual classifica o tempo de sobrevivência dos dados em ordem crescente, $t_1 \leq t_2 \leq \dots \leq t_q$, até a ocorrência do evento de interesse, e permite a representação gráfica da distribuição de probabilidade de sobrevivência. Assim, a estimativa para a duração do afastamento por acidente ou doença relacionada ao trabalho é dada por $\hat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{r_j}\right)$. A quantidade de indivíduos que saíram do período de afastamento no tempo t_j é dada por d_j , e o número de

¹Como os dados só nos permitem observar os afastamentos ocorridos no ano de 2017, a quantidade máxima de dias de afastamento do trabalho registrada na amostra é de 365 dias. Assim, pode existir uma alta quantidade de trabalhadores que não retornaram ao emprego antes de um ano, mesmo que o período de afastamento registrado tenha sido de 365 dias (quantidade máxima permitida na amostra referente ao ano de 2017). Logo, essas observações foram consideradas censuradas, as quais correspondem a 7,6% da amostra total referente à Análise II.

indivíduos em risco antes desse período é representado por r_j . Como destacado por Greene (2012), uma das principais vantagens desse método é o fato de este permitir estimar a função de sobrevivência de forma não paramétrica, além de eliminar a possibilidade de dependência de duração ao assumir independência entre os tempos de sobrevivência e que estes são identicamente distribuídos.

Outro conceito importante nos modelos de duração é a função de risco, $h(t)$, que, de acordo com Cleves et al. (2008), mede a probabilidade de que o evento de falha (fim do período de afastamento do trabalho) ocorra em um intervalo específico do tempo, depois de o sujeito ter sobrevivido até o início de tal intervalo: $h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < T < t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}$.

Existem vários tipos de modelos que podem ser usados em estudos de dados de sobrevivência, os quais geralmente são classificados em não paramétricos, paramétricos, e semi-paramétricos. Neste último grupo, inclui-se o Modelo de Regressão de Cox, o qual pressupõe que as razões de riscos são constantes ao longo do tempo. As distribuições paramétricas mais utilizadas na Análise de Sobrevivência são a Exponencial, Weibull, Log-logística e Log-normal, nas quais a distribuição do tempo de sobrevivência é especificada. Diferente das técnicas não-paramétricas, elas permitem a inclusão de diferentes covariáveis na análise empírica, bem como testar a significância estatística de cada uma delas. Ao contrário dos métodos de tempo de sobrevivência semi-paramétricos, que se concentram apenas na influência dos fatores no risco de falha, os modelos de sobrevivência paramétricos, citados acima, também permitem calcular o efeito das covariáveis sobre o tempo de sobrevivência, quando estimados na métrica *Accelerated Failure-Time* (AFT). Vale destacar também que, os modelos Exponencial e Weibull, podem ser estimados tanto por meio do modelo *Proportional Hazards* (PH) quanto pelo método AFT, enquanto as distribuições Log-logística e Log-normal são implementadas apenas por intermédio do AFT. A Tabela 1 resume as funções de sobrevivência associadas a cada modelo e a cada distribuição já mencionada.

Nos modelos paramétrico PH assume-se a proporcionalidade das taxas de riscos, e que as covariáveis têm um efeito multiplicativo na função de risco: $h(t) = h_0(t) \exp(\beta'X)$, em que β é um vetor de coeficientes de regressão, e X o conjunto de covariáveis. A função de linha de base $h_0(t)$ – o risco quando todas as covariáveis são iguais a zero – pode ser deixada sem especificação, resultando, portanto, no modelo PH de Cox, ou pode assumir uma forma paramétrica específica. Quando o modelo de duração é escrito na métrica AFT, o logaritmo natural do tempo de sobrevivência, $\log(T)$, é expresso como uma função linear das covariáveis: $\log(T) = \beta'X + \varepsilon$, e a forma da distribuição do termo de erro, ε , determina o modelo de regressão. Nesse modelo, o efeito de cada variável explicativa incide não sobre o risco de falha, como no modelo PH, mas sim sobre o tempo de sobrevivência.

Tabela 1: Função de sobrevivência das distribuições paramétricas

Distribuição	Métrica	Função de Sobrevivência	Parametrização	Parâmetros Auxiliares ¹
Exponencial	PH	$\exp(-\lambda_j t_j)$	$\lambda_j = \exp(x_j \beta)$	
Exponencial	AFT	$\exp(-\lambda_j t_j)$	$\lambda_j = \exp(-x_j \beta)$	
Weibull	PH	$\exp(-\lambda_j t_j^p)$	$\lambda_j = \exp(x_j \beta)$	p
Weibull	AFT	$\exp(-\lambda t_j^p)$	$\lambda_j = \exp(-p x_j \beta)$	p
Log-logística	AFT	$\left\{1 + (-\lambda_j t_j)^{1/\gamma}\right\}^{-1}$	$\lambda_j = \exp(-x_j \beta)$	γ
Log-normal	AFT	$1 - \Phi\left(\frac{\log(t_j) - \mu_j}{\sigma}\right)$	$\mu_j = x_j \beta$	σ

Fonte: Elaborado a partir de StataCorp (2013). Nota: ¹Parâmetros estimados a partir dos dados.

Diante do exposto, a análise da duração do período de afastamento do emprego por motivo de lesão ou doença relacionada ao trabalho no Ceará será realizada em duas etapas. Na

primeira, o Teste de Kaplan-Meier será utilizado com o intuito de facilitar a visualização das taxas de sobrevivência entre diferentes grupos de trabalhadores e firmas. Na segunda etapa, busca-se capturar o efeito de características individuais e ocupacionais do trabalhador, bem como da empresa, no tempo de ausência do emprego em decorrência de acidentes ou doenças ocupacionais, a partir da utilização de técnicas paramétricas. Mais especificamente, os modelos Exponencial, Weibull, Log-logística e Log-normal serão estimados na métrica AFT e, a partir do Critério de Informação de Akaike (AIC), um estimador projetado por Akaike (1974) para comparar a adequação dos modelos de duração aos dados, será identificado e selecionado aquele de melhor desempenho.

4 RESULTADOS

4.1 Análise Descritiva

Objetiva-se nesta pesquisa analisar o efeito de características individuais do trabalhador, de seu emprego, e da empresa, na probabilidade de afastamento da atividade laboral devido à ocorrência de acidente ou de doença ocupacional (Análise I) no estado do Ceará. Busca-se também, averiguar como esses fatores afetam a quantidade de dias de trabalho perdidos em decorrência de tais infortúnios, ao restringir a amostra para trabalhadores que sofreram lesões ou tiveram doenças ocupacionais que resultaram em afastamento de pelo menos um dia (Análise II).

As estatísticas descritivas da amostra utilizada em ambas as análises são reportadas na Tabela 2. Inicialmente, observa-se que 5.118 trabalhadores cearenses, cerca de 0,26% da primeira amostra, foram afastados do emprego por motivos de acidentes de trabalho ou de doenças geradas em virtude da atividade laborativa desempenhada, e que a duração média de tal afastamento foi de 113 dias. Verifica-se também que, 56,3% dos trabalhadores considerados na Análise I eram do sexo masculino, 16,8% se declararam brancos, e apenas 0,93% possuíam algum tipo de deficiência. A idade média desses trabalhadores é de 36 anos, os quais, em sua maioria (56,5%), tinham ensino médio completo. Quanto às características profissionais, observa-se que eles tinham, em média, 4,8 anos de experiência, pouco mais de 8% estavam no primeiro emprego, e mais da metade deles trabalhavam por mais de 40 horas semanais. Além disso, 22% dos trabalhadores presentes na amostra da Análise I estavam engajados em empregos cuja escolaridade mínima requerida era o ensino superior, e mais de 70% recebiam entre 1,01 e 3 salários-mínimos. No que diz respeito às características da empresa e dos setores de atividade econômica, as estatísticas indicam que a maioria desses trabalhadores estava empregada em empresas de grande porte (52%), e mais da metade trabalhavam no setor de serviços (58,8%).

Em relação à amostra de indivíduos que sofreram acidentes de trabalho ou que incorreram em doenças ocupacionais que resultaram em afastamento da atividade laborativa de pelo menos um dia, Análise II, as estatísticas reportadas na Tabela 2 revelam que, 74,6% dos acidentados eram homens, 14,2% e 59% eram brancos e tinham ensino médio completo, respectivamente, enquanto apenas 7,7% tinham formação de nível superior. Cerca de 2,6% dos trabalhadores portavam algum tipo de deficiência, 5,5% estavam no primeiro emprego, e eram menos experientes (3,88 anos), se comparados ao grupo de trabalhadores considerado na Análise I. Observa-se também que, pouco menos de 7% dos indivíduos que sofreram acidentes de trabalho exerciam funções que exigiam no mínimo o ensino superior completo, e apenas 17,1% deles trabalham 40 horas por semana ou menos. Quanto às características da empresa, tem-se que, 23,8% dos trabalhadores eram de empresas micro e de pequeno porte, enquanto 14,1% e 38,2%, aproximadamente, trabalhavam em médias e grandes empresas, respectivamente. Por fim, em relação aos setores econômicos, os resultados indicam que o setor

de serviços concentra o maior percentual de trabalhadores afastados do emprego por lesão ou doença ocupacional (43,7%), e que os menores percentuais foram registrados pelas atividades relacionadas aos setores de construção civil (11,1%) e de agricultura e pesca (3%).

Tabela 2: Estatísticas Descritivas

Variáveis	Análise I		Análise II	
	Média	Desvio Padrão	Média	Desvio Padrão
Variáveis Dependentes				
AcidenteTrabalho	0,00256	0,05059	-	-
DiasAfastamentoTrabalho	-	-	113,1166	108,7846
Características Individuais				
Sexo	0,56291	0,49602	0,74638	0,43512
Cor	0,16814	0,37399	0,14224	0,34933
Idade	36,3760	11,2371	37,3954	11,0667
FundIncompleto	0,09073	0,28723	0,14224	0,34933
FundCompleto	0,14095	0,34797	0,19069	0,39289
MédioCompleto	0,56528	0,49572	0,59046	0,49179
SuperiorCompleto	0,20303	0,40225	0,07659	0,26596
PortDeficiência	0,00928	0,09592	0,02598	0,15911
Características Ocupacionais				
Experiência	4,79581	7,15507	3,88272	4,89977
PrimeiroEmprego	0,08419	0,27767	0,05549	0,22895
FormMinCBO	0,22027	0,41443	0,06701	0,25007
HrsTrabalho	0,34116	0,47410	0,17096	0,37651
Renda00	0,13062	0,33698	0,12387	0,32947
Renda01	0,72518	0,44641	0,79327	0,40499
Renda02	0,07841	0,26881	0,05803	0,23382
Renda03	0,06577	0,24789	0,02481	0,15557
Características da Empresa				
MicroEmpresa	0,18954	0,39194	0,23837	0,42613
EmpresaPP	0,19800	0,39849	0,23837	0,42613
EmpresaMP	0,09298	0,29041	0,14126	0,34832
EmpresaGP	0,51945	0,49962	0,38198	0,48592
Setores de Atividade Econômica				
Agricultura_Pesca	0,01705	0,12946	0,02950	0,16923
Indústria	0,14932	0,35640	0,20750	0,40555
Comércio	0,18602	0,38912	0,21473	0,41067
Serviços	0,58825	0,49214	0,43708	0,49607
ConstCivil	0,05934	0,23627	0,11117	0,31438
Nº Observações	1.993.919		5.118	

Fonte: Elaboração própria.

4.2 Determinantes da Probabilidade de Afastamento por Acidentes ou Doenças Ocupacionais

Os resultados da análise de regressão logística são apresentados na Tabela 3. De início, verifica-se que os riscos de afastamento do trabalho devido à ocorrência de acidentes ou doenças ocupacionais são maiores para os homens do que para as mulheres, conforme já evidenciado pela literatura (BHATTACHERJEE et al., 2003; NAKATA et al., 2006). Estima-se que a probabilidade de ser afastado do emprego por motivos de acidentes ou doenças relacionadas ao trabalho aumenta em 0,11% quando o indivíduo é do sexo masculino. Em termos de razões de chances, tem-se que as chances de afastamento para os homens superam as das mulheres em 78,6%. Ademais, conforme ilustrado na Figura 1(a), observa-se que a disparidade na probabilidade de afastamento do trabalho por esses motivos entre homens e mulheres aumenta com a idade, e que os riscos de acidentes são significativamente maiores para o grupo de trabalhadores que passaram da idade de aposentadoria, sugerindo a existência

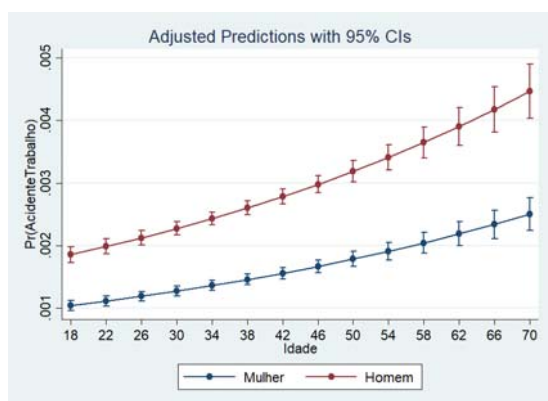
de consequências importantes do envelhecimento da força de trabalho cearense para a saúde e segurança ocupacional dos trabalhadores.

Tabela 3: Resultados da análise de regressão logística

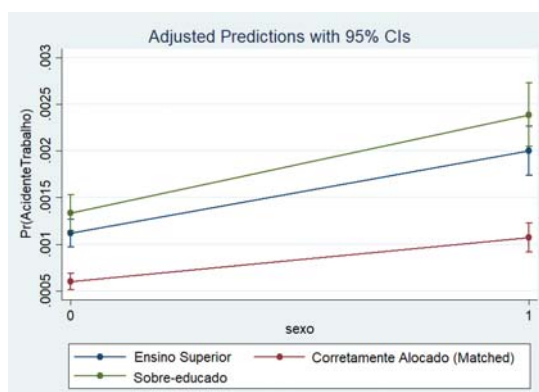
Variáveis	Coefficiente	Desvio Padrão	Odds Ratio (O.R.)	Desvio Padrão	Efeito Marginal	Desvio Padrão
Características Individuais						
Sexo	0,57994***	0,03368	1,78593	0,06016	0,00113	0,00006
Cor	-	0,04034	0,81214	0,03276	-0,00040	0,00007
Idade	0,01692***	0,00140	1,01706	0,00142	0,00003	2,75e-06
FundCompleto	-0,02581	0,04989	0,97451	0,04862	-0,00005	0,00009
MédioCompleto	-0,06050	0,04464	0,94128	0,04202	-0,00011	0,00008
SuperiorCompleto	-	0,07655	0,74389	0,05694	-0,00058	0,00014
PortDeficiência	0,81754***	0,08933	2,26492	0,20232	0,00160	0,00017
Características Ocupacionais						
Experiência	0,07632***	0,00736	1,07931	0,00795	0,00015	0,00001
Experiência2	-	0,00033	0,99634	0,00033	-7,18e-06	6,31e-07
PrimeiroEmprego	0,26055***	0,06697	1,29765	0,08691	0,00051	0,00013
FormMinCBO	-	0,06888	0,44884	0,03091	-0,00157	0,00013
HrsTrabalho	-	0,04713	0,72503	0,03417	-0,00063	0,00009
Renda00	0,31093***	0,10547	1,36470	0,14394	0,00061	0,00020
Renda01	0,27032***	0,09737	1,31038	0,12759	0,00053	0,00019
Renda02	0,41583***	0,10770	1,51563	0,16323	0,00081	0,00021
Características da Empresa						
MicroEmpresa	0,22011***	0,04074	1,24621	0,05077	0,00043	0,00007
EmpresaPP	0,19941***	0,03980	1,22068	0,04859	0,00039	0,00007
EmpresaMP	0,39376***	0,04602	1,48254	0,06823	0,00077	0,00009
Setores de Atividade						
Econômica						
Agricultura_Pesca	-0,03620	0,09294	0,96444	0,08964	-0,00007	0,00018
Indústria	-	0,05539	0,85145	0,04716	-0,00031	0,00010
Comércio	-	0,05514	0,76568	0,04222	-0,00052	0,00010
Serviços	-	0,05234	0,72637	0,03801	-0,00062	0,00010
Constante	-	0,13388	0,00091	0,00012	-	-
N° Observações						1.993.919

Fonte: Elaboração própria. Nota: * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$.

Figura 1: Lacunas de gênero



(a) Influência da idade na probabilidade de afastamento por acidente ou doença ocupacional entre homens e mulheres



(b) Probabilidade de afastamento por acidente ou doença ocupacional de acordo com as condições educacionais e ocupacionais dos trabalhadores graduados

As conclusões a respeito da relação entre idade e acidentes de trabalho geralmente têm sido contraditórias. Inúmeras pesquisas apontam uma prevalência de acidentes de trabalho maior para trabalhadores mais jovens (NAKATA et al., 2006; ADERAW et al., 2011; RIOS et al., 2015). Outras, porém, sugerem uma relação positiva entre o envelhecimento da força de trabalho e lesões ocupacionais (HASEBE; SAKAI, 2018), e que os trabalhadores mais velhos têm uma probabilidade maior de sofrer acidentes mais graves (BANDE; LÓPEZ-MOURELO, 2015). Os resultados desta pesquisa corroboram estes últimos achados. As estimativas, presentes na Tabela 3, indicam que um ano a mais de idade aumenta as chances de afastamento por acidente ou doença em aproximadamente 2%. Como destacado por Bande e López-Mourelo (2015), os trabalhadores mais velhos, por um lado, geralmente são mais experientes que os mais jovens. Isso tende a reduzir os riscos de lesões no ambiente de trabalho. No entanto, enfrentam uma redução em sua capacidade de evitar acidentes inesperados, além de estarem mais propensos a sofrerem de diminuições visuais e auditivas, bem como de estarem mais confiantes na condução de suas atividades em razão de sua maior experiência e, com isso, podem desconsiderar medidas de prevenção importantes para certos riscos, elevando as chances de acidentes.

Ainda em relação às características individuais, os resultados revelam que as chances de afastamento por lesão ou doença ocupacional são significativamente menores para trabalhadores brancos, e maiores para os que possuem algum tipo de deficiência. As chances de afastamento para as pessoas brancas são 18,8% menores que as de pessoas não brancas. Quanto aos trabalhadores com deficiência, esses são duas vezes mais propensos ao afastamento quando comparados aos não deficientes.

Quando se analisa a relação entre educação e a probabilidade de afastamento do trabalho em decorrência de acidentes ou de doenças ocupacionais, nota-se que, apenas o coeficiente associado ao grau de escolarização de nível superior é estatisticamente significativo. De acordo com as estimativas do efeito marginal, verifica-se que, ter um diploma de graduação reduz a probabilidade de afastamento por acidentes de trabalho em aproximadamente 0,06%. Resultados similares foram encontrados por Cavalcante (2013). Em um estudo para a região Nordeste do Brasil, o autor observa que os indivíduos com ensino superior possuem 19,3% a menos de chances de sofrerem acidentes de trabalho quando comparados com seus pares de menor instrução. Tais resultados podem ser explicados pelas melhores oportunidades de aquisição de conhecimento e de habilidades de trabalhadores mais instruídos, o que pode aumentar a presença de comportamentos mais seguros no ambiente de trabalho. Torna-se evidente a importância da educação para a prevenção de acidentes ocupacionais. No entanto, ainda não está claro se a educação de nível superior atua como um fator de proteção mesmo quando os graduados são sobre-educados, ou seja, trabalham em ocupações que exigem menores níveis de qualificações, aquém do ensino superior.

Nas últimas décadas, o sistema de ensino superior brasileiro apresentou um importante processo de expansão decorrente principalmente do desenvolvimento e implantação de programas educacionais voltadas para a ampliação do acesso ao ensino universitário². Pesquisas mostram que essa expansão contribuiu não somente para o crescimento da escolaridade média da população brasileira, mas também para o aumento do nível de renda, salário, e taxas ocupacionais (ROCHA et al., 2017). Apesar desse avanço, a demanda por esses profissionais de maior qualificação pelo mercado de trabalho parece não ter crescido no mesmo ritmo que a expansão de sua oferta (PAULI et al., 2012), criando-se, assim, um ambiente

² Fundo de Financiamento Estudantil (FIES), e o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades (REUNI).

propício ao surgimento de trabalhadores em situação de sobre-educação, que aceitam trabalho em ocupações que exigem aquém de seu nível de escolaridade.

Os efeitos desse fenômeno sobre os retornos econômicos (REIS, 2015; ANNEGUES; SOUZA, 2020), a mobilidade ocupacional (BAERT et al., 2013; ACOSTA-BALLESTEROS et al., 2018), a produtividade (TSANG, 1987) e a satisfação no emprego dos trabalhadores (ALLEN; VAN DER VELDEN, 2001), têm sido tradicionalmente investigados pela literatura. No entanto, um campo de pesquisa ainda escasso, mas que também merece atenção é o que busca estabelecer uma relação entre incompatibilidade de educação e acidentes de trabalho.

Com o intuito de acrescentar a essa literatura, calculou-se o efeito de graduados aceitarem trabalho em ocupações que exigem abaixo de seu nível de escolaridade (sobre-educados) na probabilidade de serem afastados do trabalho por acidentes ou doenças ocupacionais. Tais resultados foram obtidos a partir da estimativa do efeito marginal das variáveis de escolaridade e aquela relacionada ao nível de educação requerido pela ocupação (FormMinCBO) na probabilidade de acidentes, mantendo as demais variáveis constantes em seus valores médios. Tais resultados são apresentados graficamente na Figura 1(b). As linhas em vermelho e em verde representam, respectivamente, a probabilidade dos graduados adequadamente alocados (engajados em empregos condizentes com o seu nível de escolaridade) e sobre-educados, serem afastados do trabalho em decorrência de acidentes ou doenças ocupacionais. A linha em azul, por outro lado, representa a probabilidade do graduado ser afastado do emprego pelos mesmos motivos, independente do nível de escolaridade exigido pela ocupação. Todas essas estimativas foram estratificadas por gênero a fim de capturar as disparidades existentes entre homens (sexo=1) e mulheres (sexo=0).

A partir da Figura 1(b), observa-se que a probabilidade do trabalhador adequadamente alocado no mercado de trabalho, ou seja, com educação condizente com a que é requerida pela ocupação, ser afastado devido a acidentes ou doenças ocupacionais é estimada em 0,11%. No entanto, quando o graduado aceita emprego em atividades que exigem aquém de sua formação, essa probabilidade salta para 0,24%. Nota-se, portanto, que a probabilidade de acidentes dobra quando os trabalhadores são sobre-educados, em relação aqueles adequadamente ocupados, e o mesmo se verifica para as mulheres. A partir desses resultados, pode-se concluir que os graduados sobre-educados são significativamente mais propensos a se envolverem em acidentes de trabalho do que seus pares que se encontram engajados em empregos condizentes. O resultado vai ao encontro de Premji e Smith (2013), e possibilita um maior debate sobre acidente de trabalho voltado especificamente para o fator de incompatibilidade educacional no mercado ocupacional como um possível determinante. Como enfatizado por Premji e Smith (2013), os trabalhadores com escolaridade em excesso podem subestimar o conhecimento e nível de experiência necessário para a realização de suas atividades laborativas com segurança, além de estarem expostos a menores oportunidades de treinamentos no emprego, uma vez que eles tendem a substituir educação por treinamento. Além disso, é possível que os trabalhadores sobre-educados estejam mais inclinados a aceitarem uma carga horária de trabalho mais elevada, ou até mesmo de aceitar um emprego secundário, seja no setor formal ou informal, a fim de compensar a penalização de renda incorrida ao aceitar emprego em ocupações que exigem abaixo de suas qualificações. Isso poderá contribuir para o aumento do cansaço e estresse devido ao aumento das horas trabalhadas e, conseqüentemente, para o aumento da probabilidade de acidentes.

No que diz respeito às características profissionais dos trabalhadores, verifica-se que, maiores jornadas de trabalho estão associadas a maior probabilidade de afastamento. As pessoas que trabalhavam 40 horas semanais ou menos, são menos propensas a serem afastadas por causa de acidentes do que aquelas que trabalhavam por mais tempo (OR=0,725). Resultados análogos são evidenciados por Tadesse e Kumie (2007) e Dong et al. (2015), os quais observaram que a exposição a longas horas de trabalho aumenta as chances de lesões ocupacionais. Os

funcionários que trabalham por um longo período podem sofrer fadiga ou estresse que além de reduzir a produtividade, tornam-os mais propensos a erros e acidentes (PENCAVEL, 2015; SOARES et al., 2018). Os resultados também revelam que os indivíduos que atuam em ocupações que exigem maiores níveis de qualificações, no caso, ensino superior completo, bem como aqueles que detêm um maior nível de renda, têm menor probabilidade de serem afastados do trabalho por motivo de acidente ou doença adquirida no ambiente de trabalho, o que, por sua vez, pode indicar que menores salários estejam associados a atividades de alto risco (LIMA et al., 1999).

Em contraste com alguns estudos referentes ao tema (NAKATA et al., 2006; ADERAW et al., 2011), esta pesquisa mostra que os riscos de afastamento por acidentes ou doenças relacionadas ao trabalho aumentam com o tempo de serviço na empresa, mas que crescem a taxas decrescentes. Resultados análogos foram encontrados por Keyserling (1983), Cavalcante (2013) e Belchior (2018), os quais também confirmam a existência de uma relação não linear, em forma de U invertido, entre experiência no mesmo emprego e a probabilidade de sofrer acidentes de trabalho, sugerindo que os riscos de acidentes crescem com o tempo de serviço, mas que, a partir de certo ponto, entra em declínio. Essa relação positiva pode estar associada ao fato de os trabalhadores se tornarem mais confiantes com o manuseio de materiais e execução de suas tarefas de trabalho à medida que se ganha mais experiência, o que pode contribuir para a ocorrência de descuidos e, conseqüentemente, de acidentes. Ademais, espera-se que o treinamento e a maior supervisão dos novos trabalhadores, que entraram recentemente na empresa, contribuam para a redução dos riscos de acidentes nesse grupo, tornando-os menos propensos a sofrerem acidentes em relação aos que possuem mais tempo de serviço.

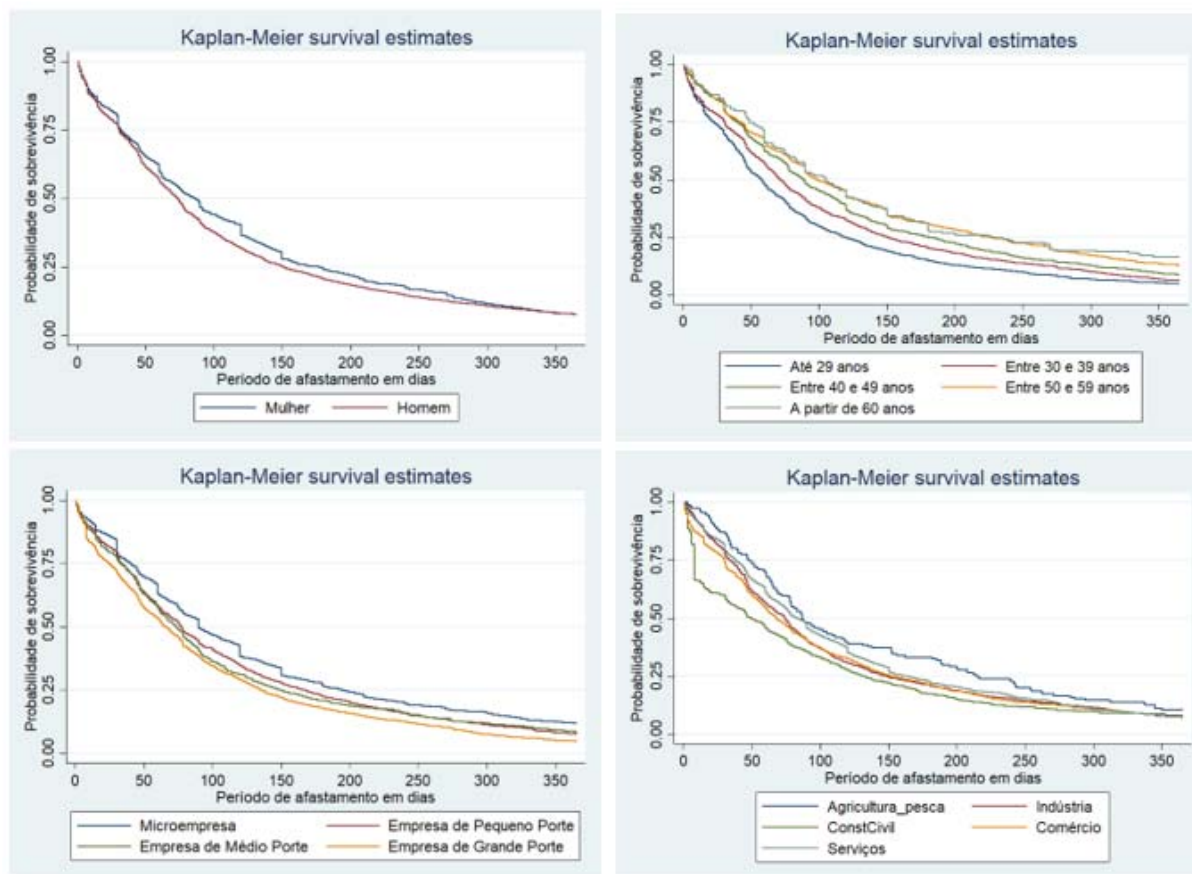
Em relação às características do estabelecimento, constata-se que a probabilidade de acidentes ou de doenças é relativamente menor para empresas de grande porte. As chances de afastamento por essas causas em empresas de micro, pequeno, e de médio porte, superam as de empresas de grande porte em aproximadamente 25%, 22%, e 48%, respectivamente. De acordo com López-Arquillos e Rubio-Romero (2016), a dificuldade de controle dos riscos de lesões e de acidentes enfrentada pelas empresas de menor porte, em comparação com as de maior porte, em virtude da limitação de recursos materiais e humanos, pode ser uma das explicações possíveis para esses resultados. Além disso, como destacado por Fabiano et al. (2004), as empresas de maior porte normalmente contam com uma gestão das questões de prevenção mais sólida, e podem possuir melhores atividades de treinamento e informações em relação às empresas menores. Por fim, as estimativas para os setores de atividade econômica mostram que as chances de acidentes de trabalho são significativamente maiores na construção civil do que na indústria, comércio e serviços, reforçando os resultados de Cavalcante (2013), o qual, em um estudo para a região Nordeste do Brasil, encontrou evidências de que os trabalhadores da construção civil enfrentam maiores probabilidades de acidentes do que seus pares empregados nos demais setores. Estima-se que as chances de afastamento por acidentes ou doenças na indústria e no comércio são, respectivamente, 14,9% e 23,4% menores do que na construção civil. Já para o setor de serviços, estima-se uma redução de 27,4%.

4.3. Determinantes da Duração do Período de Afastamento por Acidentes ou Doenças Ocupacionais

As estimativas das funções de sobrevivência, obtidas a partir do estimador de Kaplan-Meier, para algumas das principais variáveis utilizadas na estimação dos modelos de duração, são apresentadas na Figura 2. Elas mostram a distribuição de probabilidade da continuidade da ausência dos trabalhadores de suas atividades laborais por motivos de acidentes ou doenças relacionadas ao emprego. De início, observa-se que os trabalhadores do sexo masculino, em relação ao feminino, possuem maiores chances de retornarem às atividades de trabalho após serem afastados devido a acidentes ou doenças ocupacionais. No centésimo dia de afastamento,

por exemplo, a chance dos homens permanecerem afastados do emprego por esses mesmos motivos é de aproximadamente 38%, enquanto que para as mulheres a perspectiva é de 44%. Todavia, essa disparidade diminui com o tempo e praticamente desaparece após um intervalo de 300 dias. Resultados semelhantes são encontrados por Fontaneda *et al.* (2019), os quais também evidenciam um período de afastamento maior para as mulheres.

Figura 2: Probabilidade de sobrevivência, estimador de Kaplan-Meier



Fonte: Elaboração própria.

Em relação à idade, verifica-se que, durante todo o período de análise, a probabilidade da duração do tempo de afastamento por doença ou acidente é claramente maior para os trabalhadores mais velhos do que para os mais jovens, sugerindo que os trabalhadores de idade mais avançada não são apenas mais propensos a se envolverem em acidentes, mas também permanecem em licença médica por um período de tempo significativamente maior, corroborando os achados de Bande e Lopéz-Mourelo (2015).

Além de apresentarem as maiores taxas de contratações, menor probabilidade de ocorrência de afastamentos devido a acidentes ou doenças ocupacionais, as empresas de grande porte são também as que registram menores períodos de ausência do trabalho, ou seja, maiores taxas de retorno ao emprego após a ocorrência de acidentes ou doenças que resultaram em afastamento de pelo menos um dia. Nas micro empresas, por outro lado, a duração do tempo de afastamento é significativamente maior, e essa desvantagem se mantém durante todo o período de estudo. Diferenças nas taxas de sobrevivência entre os estabelecimentos de pequeno e médio porte aparecem após os primeiros 60 dias de afastamento do trabalho, mas desaparecem logo em seguida, após um período de 200 dias.

Como visto anteriormente, a probabilidade de os trabalhadores da construção civil serem afastados de suas atividades laborais por doença ou acidentes de trabalho supera, de forma significativa, a de trabalhadores ocupados na indústria, comércio e serviços. No entanto,

os resultados apresentados na Figura 2 indicam que os dias de trabalho perdidos parecem menores na construção civil do que nos demais setores, mas que essa diferença entra em declínio conforme se aumenta o tempo de ausência. Por outro lado, a duração do período de afastamento é significativamente maior nos setores agropecuário e de pesca. Em relação às atividades relacionadas à indústria e comércio, nota-se que suas funções de sobrevivência seguem o mesmo padrão, e que as diferenças em suas taxas de sobrevivência encontram-se praticamente ausentes em vários momentos do tempo.

Com o auxílio de modelos de duração paramétricos, foi possível verificar se essas diferenças permanecem mesmo após o controle de um conjunto de características individuais e ocupacionais dos trabalhadores. O modelo AFT foi estimado considerando todas as distribuições mencionadas anteriormente: Exponencial, Weibull, Log-logística e Log-normal. Por meio do Critério de Informação de Akaike (AIC), cujos resultados são apresentados na Tabela 4, foi possível identificar a distribuição que melhor se adequa aos dados. Optou-se pela distribuição Weibull, por ter registrado o menor valor do AIC e, portanto, o melhor desempenho.

Tabela 4: Critério de Informação de Akaike (AIC)

Distribuição	Exponencial	Weibull	Log-logística	Log-normal
AIC	16907,22	16872,35	17123,19	17287,23

Fonte: Elaboração própria.

Apresentam-se na Tabela 5 as estimativas dos coeficientes do modelo Weibull AFT, expressos em Time Ratio. Os resultados indicam que, mesmo após o controle de um conjunto de características individuais e ocupacionais dos trabalhadores, a quantidade de dias de trabalho perdidos por motivos de doença ou de acidentes no local de trabalho são maiores para as mulheres. O período de afastamento do trabalho por essas causas reduz em 7,6% quando o indivíduo é do sexo masculino, estando em conformidade com os resultados encontrados anteriormente. Portanto, embora as mulheres tenham menor probabilidade de sofrerem acidentes ou doenças ocupacionais, a velocidade de seu retorno ao trabalho após serem afastadas é menor que a dos homens. Em relação à idade, mais uma vez as estimativas sugerem que a duração do período de afastamento é significativamente maior para trabalhadores mais velhos do que para os mais jovens, sugerindo, portanto, que a gravidade dos acidentes, medida pela quantidade de dias de trabalho perdidos, tende a ser maior para aqueles de idade mais avançada, os quais necessitam de um período de tempo mais elevado para a recuperação. Esse resultado condiz com o de Bande e Lopéz-Moureló (2015), os quais, a partir da análise de modelos de duração, chegaram à mesma conclusão. Quanto à escolaridade, verifica-se que a continuidade do tempo de afastamento é 21,7% menor para trabalhadores com ensino superior, em relação aqueles sem instrução ou que não concluíram o ensino fundamental.

Tabela 5: Estimativas paramétricas com distribuição Weibull

Variáveis	Time Ratio	Desvio Padrão	Variáveis	Time Ratio	Desvio Padrão
Características Individuais			Renda01	1,55470***	0,17733
Sexo	0,92360**	0,03679	Renda02	1,46972***	0,18411
Cor	1,01087	0,04625	Características da Empresa		
Idade	1,01562***	0,00168	MicroEmpresa	1,69201***	0,07797
FundCompleto	1,13245**	0,06451	EmpresaPP	1,33361***	0,05658
MédioCompleto	1,01889	0,05162	EmpresaMP	1,34883***	0,06760
SuperiorCompleto	0,78284**	0,07739	Setores de Atividade Econômica		
PortDeficiência	1,05797	0,10542	Agricultura_Pesca	1,67010***	0,17607
Características Ocupacionais			Indústria	1,06922	0,06595
Experiência	1,12536***	0,00927	Comércio	1,12394*	0,06927

Experiência2	0,99624***	0,00027	Serviços	1,35212***	0,07947
PrimeiroEmprego	0,75066***	0,06567	Constante	20,44215***	3,17771
FormMinCBO	1,22389**	0,11476	<i>ln_p</i>	-0,07035***	0,01177
HrsTrabalho	0,87439**	0,04773	<i>p</i>	0,93206***	0,01097
Renda00	2,19605***	0,26891	Nº Observações		5.118

Fonte: Elaboração própria. Nota: * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$.

Em relação às características ocupacionais, os resultados indicam que o tempo de afastamento para trabalhadores de primeiro emprego e para os que trabalhavam 40 horas semanais ou menos, é 25% e 12,6% menor que o de seus pares, respectivamente. Além disso, observa-se que a duração do período de ausência do trabalho é maior para os profissionais que recebiam menores salários. Como os baixos rendimentos podem estar associados a atividades de alto risco (LIMA et al., 1999), é possível que a gravidade dos acidentes também seja maior nessas atividades, havendo a necessidade de um período de afastamento mais elevado para a recuperação dos funcionários acidentados. O coeficiente da variável referente à experiência se mostrou positivo, sugerindo que a duração do afastamento aumenta com o tempo de serviço no mesmo emprego, porém, a taxas decrescentes. A mesma relação foi encontrada por Belchior (2018), em um estudo recente para o Brasil. A partir de dados da RAIS de 2013, e da estimação de um modelo de regressão Tobit, o autor encontra que, em média, um ano a mais de experiência no mesmo emprego aumenta o período de afastamento em aproximadamente 10 dias.

Para as variáveis referentes ao tamanho do estabelecimento, as estimativas sugerem que, além de a probabilidade de acidente ser menor para trabalhadores empregados em empresas de grande porte, os acidentes sofridos tendem a ser menos graves. Estima-se, por exemplo, que o tempo de ausência dos trabalhadores que foram afastados devido a acidentes ou doenças ocupacionais é quase que duas vezes maiores nas microempresas, e cerca de 34% maior nos estabelecimentos de pequeno e médio porte, do que nas empresas de grande porte. Fabiano et al. (2004), destacam que as empresas maiores normalmente contam com uma gestão das questões de prevenção de acidentes mais sólida, e podem ter melhores atividades de treinamento e informações em relação as empresas de menor porte. Isso pode contribuir não somente para a redução dos acidentes de trabalho em firmas maiores, mas também para a redução de sua gravidade. Quanto aos principais setores de atividade econômica, pode-se constatar que, embora o setor de construção civil apresente maior probabilidade de acidentes, os dias de trabalho perdidos devido a esses infortúnios parecem ser menores quando comparado aos setores agropecuário e de pesca e de serviços. Estima-se que a duração do período de afastamento do trabalho é 1,67 e 1,35 vezes maior nos setores de agricultura e pesca e de serviços, respectivamente, do que na construção civil.

5 CONSIDERAÇÕES FINAIS

A problemática que envolve acidentes de trabalho faz parte do cenário produtivo do Ceará, e caracteriza-se por ter causas variadas tanto provenientes de características do indivíduo, quanto do local de trabalho e da relação entre ambos, além de implicações que atingem de forma ampla a economia, as famílias dos acidentados e a sociedade como um todo.

Com a expectativa de melhor identificar grupos que apresentam maior propensão a sofrer algum tipo de acidente de trabalho ou desenvolver alguma patologia em decorrência de atividade laborativa, este estudo buscou identificar quais os determinantes da ocorrência de afastamento do emprego por motivos de doenças ocupacionais ou de acidentes ocorridos no ambiente de trabalho no estado cearense, e quais fatores contribuem para o período de afastamento em decorrência do evento de sinistro. Para isso, foram usados dados da RAIS de 2017 que contêm informações acerca dos indivíduos e das empresas onde trabalham, e como estratégia metodológica, foram estimados um modelo de regressão logística e uma análise de sobrevivência.

Em relação aos achados da regressão logística, foi verificado que ser do sexo masculino aumenta as chances de afastamento por acidente ou doenças ocupacionais e esse resultado é acentuado conforme os anos de vida aumentam, principalmente ao ultrapassar a idade de aposentadoria. As estimativas sugerem ainda que as chances de afastamento ampliam-se conforme a idade aumenta, o que corrobora a literatura que sustenta essa evidência (BANDE; LÓPEZ-MOURELO, 2015). Sobre as demais características do indivíduo, se este for branco, as chances de ser afastado em decorrência de acidente ou doença reduzem, enquanto se o trabalhador possuir algum tipo de deficiência, essas chances aumentam. Ao observarmos os resultados que refletem o nível de escolaridade, vemos que ter título de graduação reduz as chances de ocorrência do evento, ao passo que ser sobre-educado aumenta as chances de afastamento. Quanto às características ocupacionais, constatou-se que longas jornadas de trabalho (superior a 40h/semana) aumentam os riscos de o trabalhador ser afastado, e que quanto maior o tempo de experiência na empresa, maiores as chances de ocorrência do evento, apesar de esse crescimento evoluir a taxas decrescentes. Em relação às características do estabelecimento, se a empresa for de grande porte, a probabilidade de um profissional ser afastado por causa de acidente de trabalho é reduzida. E no que se refere aos setores, trabalhar na construção civil aumenta as chances de afastamento em relação aos demais.

Os resultados da análise de sobrevivência evidenciam que a duração do período de afastamento em decorrência de doenças ocupacionais ou de acidentes ocorridos no ambiente de trabalho é maior para trabalhadores do sexo feminino, de idade mais avançada, e de menor instrução. Em relação às características ocupacionais, verificou-se que o período de ausência é mais curto para o indivíduo em seu primeiro emprego e para trabalhadores que exercem 40h/semana ou menos de atividades laborais. Além disso, para os profissionais que recebem menores salários e os que possuem maior grau de experiência na função, o total de dias de trabalho perdidos é maior. Quanto ao tamanho da firma, quanto menor for o porte, maior o tempo de afastamento. E, por fim, o período em que o indivíduo passa afastado das atividades é menor para trabalhadores da construção civil quando comparados aos que se encontram empregados nos demais setores.

O conjunto de resultados apresentados neste estudo para o estado do Ceará contribui significativamente para a literatura de acidentes de trabalho tendo enfoque maior em uma de suas consequências, o afastamento. Essa ausência reflete em perdas e prejuízos para a economia e sociedade como um todo por meio de diferentes vias (gastos com saúde, previdência, perda de produtividade, etc). Como dito anteriormente, é possível evitar acidentes de trabalho, e as evidências aqui produzidas reforçam os pontos que merecem maior atenção e abre possibilidades para resoluções efetivas. Por exemplo, a realização de treinamentos para segurança e maior supervisão, principalmente, de pessoas com maior experiência e de idade mais avançada e indivíduos do sexo masculino. Além de programas que sejam voltados ao problema de sobre-educação focados em reajustar o desequilíbrio entre oferta e demanda desse público, e um maior apoio para empresas de menor porte a fim de mitigar essa problemática.

REFERÊNCIAS

ACOSTA-BALLESTEROS, Juan; DEL PILAR OSORNO-DEL ROSAL, María; RODRÍGUEZ-RODRÍGUEZ, Olga María. Overeducation of Young Workers in Spain: How Much Does the First Job Matter? Social Indicators Research. **Social Indicators Research**, v. 138, n. 1, p. 109-139, 2018.

ADERAW, Zewdie; ENGDAW, Dagnew; TADESSE, Takele. Determinants of occupational injury: a case control study among textile factory workers in Amhara Regional State, Ethiopia. **Journal of tropical medicine**, v. 2011, 2011.

AHN, Y. S.; BENA, J. F.; BAILER, A. J. Comparison of unintentional fatal occupational injuries in the Republic of Korea and the United States. **Injury prevention**, v. 10, n. 4, p. 199-205, 2004.

AKAIKE, Hirotugu. A new look at the statistical model identification. **IEEE transactions on automatic control**, v. 19, n. 6, p. 716-723, 1974.

ALLEN, Jim; VAN DER VELDEN, Rolf. Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. **Oxford economic papers**, v. 53, n. 3, p. 434-452, 2001.

ANNEGUES, Ana Cláudia; SOUZA, Wallace Patrick Santos de Farias. Retorno Salarial do Overeducation: Viés de Seleção ou Penalização ao Excesso de Escolaridade?. **Revista Brasileira de Economia**, v. 74, n. 2, p. 119-138, 2020.

BAERT, Stijn; COCKX, Bart; VERHAEST, Dieter. Overeducation at the start of the career: stepping stone or trap?. **Labour Economics**, v. 25, p. 123-140, 2013.

BANDE, Roberto; LÓPEZ-MOURELO, Elva. The impact of worker's age on the consequences of occupational accidents: empirical evidence using Spanish data. **Journal of Labor Research**, v. 36, n. 2, p. 129-174, 2015.

BELCHIOR, Carlos Alberto. A terceirização precariza as relações de trabalho? O impacto sobre acidentes e doenças. **Revista Brasileira de Economia**, v. 72, n. 1, p. 41-60, 2018.

BENA, Antonella et al. Effectiveness of the training program for workers at construction sites of the high-speed railway line between Torino and Novara: Impact on injury rates. **American journal of industrial medicine**, v. 52, n. 12, p. 965-972, 2009.

BHATTACHERJEE, Ashis et al. Relationships of job and some individual characteristics to occupational injuries in employed people: A community-based study. **Journal of occupational health**, v. 45, n. 6, p. 382-391, 2003.

CAVALCANTE, Rodrigo Ribeiro. Acidentes de trabalho: Uma análise do acidente de trabalho na construção civil na região nordeste, para o ano de 2011. 2013.

CIONI, Martina; SAVIOLI, Marco. Safety at the workplace: accidents and illnesses. **Work, employment and society**, v. 30, n. 5, p. 858-875, 2016.

CLEVES, Mario et al. **An introduction to survival analysis using Stata**. Stata press, 2008.

DONG, Xiuwen Sue; WANG, Xuanwen; LARGAY, Julie A. Occupational and non-occupational factors associated with work-related injuries among construction workers in the

- USA. **International journal of occupational and environmental health**, v. 21, n. 2, p. 142-150, 2015.
- FABIANO, Bruno; CURRÒ, Fabio; PASTORINO, Renato. A study of the relationship between occupational injuries and firm size and type in the Italian industry. **Safety science**, v. 42, n. 7, p. 587-600, 2004.
- FONTANEDA, Ignacio et al. Gender differences in lost work days due to occupational accidents. **Safety science**, v. 114, p. 23-29, 2019.
- GARCÍA-HERRERO, Susana et al. Working conditions, psychological/physical symptoms and occupational accidents. Bayesian network models. **Safety science**, v. 50, n. 9, p. 1760-1774, 2012.
- GEBREMESKEL, Teferi Gebru; YIMER, Tesfaye. Prevalence of occupational injury and associated factors among building construction workers in Dessie town, Northeast Ethiopia; 2018. **BMC research notes**, v. 12, n. 1, p. 481, 2019.
- GONÇALVES, Steffani Bez Batti; SAKAE, Thiago Mamoru; MAGAJEWSKI, Flavio Liberali. Prevalence and factors associated with work accidents in a metal-mechanic company. **Revista Brasileira de Medicina do Trabalho**, v. 16, n. 1, p. 26, 2018.
- GREENE, W. *Econometric analysis*. 7.ed. New Jersey: Pearson Prentice Hall, 2012.
- GUIMARÃES, L. A. M.; GRUBITS, S. **Série saúde mental e trabalho**. Casa do Psicólogo, 2004.
- HASEBE, Takuya; SAKAI, Tadashi. Are elderly workers more likely to die in occupational accidents? Evidence from both industry-aggregated data and administrative individual-level data in Japan. **Japan and the World Economy**, v. 48, p. 79-89, 2018.
- HOLTE, Kari Anne; KJESTVEIT, Kari; LIPSCOMB, Hester J. Company size and differences in injury prevalence among apprentices in building and construction in Norway. **Safety Science**, v. 71, p. 205-212, 2015.
- INTERNATIONAL LABOUR ORGANIZATION. **Safety in numbers**: pointers for a global safety culture at work. Geneva: International Labour Office, 2003. 33 p.
- JAFARI, Mohammad Javad et al. Relationships between certain individual characteristics and occupational accidents. **International Journal of Occupational Safety and Ergonomics**, v. 25, n. 1, p. 61-65, 2019.
- KAPLAN, Edward L.; MEIER, Paul. Nonparametric estimation from incomplete observations. **Journal of the American statistical association**, v. 53, n. 282, p. 457-481, 1958.
- KEYSERLING, W. Monroe. Occupational injuries and work experience. **Journal of safety research**, v. 14, n. 1, p. 37-42, 1983.
- LIMA, Rosângela da Costa et al. Associação entre as características individuais e sócio-econômicas e os acidentes do trabalho em Pelotas, Rio Grande do Sul, Brasil. **Cadernos de Saúde Pública**, v. 15, n. 3, p. 569-580, 1999.

LÓPEZ-ARQUILLOS, Antonio; RUBIO-ROMERO, Juan Carlos. Analysis of workplace accidents in automotive repair workshops in Spain. **Safety and health at work**, v. 7, n. 3, p. 231-236, 2016.

MINISTÉRIO DA FAZENDA. **Anuário Estatístico de Acidentes do Trabalho 2017**. Disponível em: <<http://sa.previdencia.gov.br/site/2018/09/AEAT-2017.pdf>>. Acesso em: 30 out. 2019.

NAKATA, Akinori et al. The prevalence and correlates of occupational injuries in small-scale manufacturing enterprises. **Journal of occupational health**, v. 48, n. 5, p. 366-376, 2006.

OH, Joong-Hwan; SHIN, Eui Hang. Inequalities in nonfatal work injury: the significance of race, human capital, and occupations. **Social science & medicine**, v. 57, n. 11, p. 2173-2182, 2003.

PAULI, Rafael Camargo de; NAKABASHI, Luciano; SAMPAIO, Armando Vaz. Mudança estrutural e mercado de trabalho no Brasil. **Brazilian Journal of Political Economy**, v. 32, n. 3, p. 459-478, 2012.

PENCAVEL, John. The productivity of working hours. **The Economic Journal**, v. 125, n. 589, p. 2052-2076, 2015.

PIHA, Kustaa et al. Socio-economic and occupational determinants of work injury absence. **The European Journal of Public Health**, v. 23, n. 4, p. 693-698, 2013.

PREMJI, Stephanie; SMITH, Peter M. Education-to-job mismatch and the risk of work injury. **Injury prevention**, v. 19, n. 2, p. 106-111, 2013.

REIS, M. Overeducation, undereducation and labor earnings in Brazil: a panel data analysis, Rede de Economia Aplicada, **Instituto de Pesquisa Economica Aplicada**, Working Paper 87, 2015.

REIS, Mauricio. Measuring the mismatch between field of study and occupation using a task-based approach. **Journal for labour market research**, v. 52, n. 1, p. 9, 2018.

RIOS, Marcela Andrade et al. Fatores associados a acidentes de trabalho envolvendo trabalhadores informais do comércio. **Cadernos de Saúde Pública**, v. 31, p. 1199-1212, 2015.

ROCHA, Roberto Hsu et al. A relação entre o ensino superior público e privado e a renda e emprego nos municípios brasileiros. **Pesquisa e Planejamento Econômico**, v. 47, n. 3, 2017

ROMMEL, Alexander et al. Occupational injuries in Germany: Population-wide national survey data emphasize the importance of work-related factors. **PloS one**, v. 11, n. 2, p. e0148798, 2016.

SMARTLAB. **Observatório de Segurança e Saúde no Trabalho**. Disponível em: <https://smartlabbr.org/sst>. Acesso em: 26 nov. 2019.

SOARES, Simarly M. et al. Workplace accidents in Brazil: analysis of physical and psychosocial stress and health-related factors. **RAM. Revista de Administração Mackenzie**, v. 19, n. 3, 2018.

STATAcorp. “Stata: Release 13”. Statistical Software. College Station, TX: StataCorp LP, 2013. <http://www.stata.com/manuals13/st.pdf>.

TADESSE, Takele; KUMIE, Abera. Prevalence and factors affecting work-related injury among workers engaged in Small and Medium-scale industries in Gondar wereda, North Gondor zone, Amhara Regional State, Ethiopia. **Ethiopian journal of health development**, v. 21, n. 1, p. 25-34, 2007.

TSANG, Mun Chiu. The impact of underutilization of education on productivity: A case study of the US Bell companies. **Economics of education review**, v. 6, n. 3, p. 239-254, 1987.

ANÁLISE DE CONVERGÊNCIA ESPACIAL DE PRODUTIVIDADE AGRÍCOLA APLICADA À REGIÃO NORDESTE DO BRASIL

Winnie Moreira Albuquerque*

Roberio Telmo Campos**

Vitor Hugo Miro Couto Silva***

RESUMO

Palavras-chave: Produtividade agrícola. β -convergência espacial. Nordeste brasileiro.

ABSTRACT

This research proposes an analysis of agricultural productivity for A presente pesquisa propõe uma análise da produtividade agrícola para as microrregiões geográficas do Nordeste brasileiro referente ao período de 1996 a 2017. Para a obtenção das respostas em face das perguntas aqui apresentadas, são aplicados os métodos de análise espacial e econometria espacial. De particular relevância para a teoria do crescimento e desenvolvimento econômico, a hipótese de convergência caracteriza um processo segundo o qual, inicialmente, valores de certa variável são desiguais entre países, estados ou regiões, mas essa desigualdade se reduz ao longo do tempo. Tal análise identifica, por sua vez, se há evidências para a existência de β -convergência absoluta e condicional da produtividade agrícola, controlando-se explicitamente para efeitos espaciais por subperíodos. Os resultados empíricos indicaram a presença de β -convergência absoluta e de β -convergência condicional para o Nordeste.

the geographical microregions of Northeastern Brazil for the period from 1996 to 2017. To obtain the answers to the questions presented here, the spatial analysis methods are applied and spatial econometrics. Of particular relevance to the theory of growth and economic development, the convergence hypothesis characterizes a process according to which, initially, values of a certain variable are unequal between countries, states or regions, but this inequality is reduced over time. Such analysis identifies, in turn, if there is evidence for the existence of absolute and conditional convergence of agricultural productivity, explicitly controlling for spatial effects by subperiods. The empirical results indicated the presence of absolute β -convergence and conditional β -convergence for the Northeast.

Key words: *Agricultural productivity. β -spatial convergence. Brazilian Northeast.*

Área 3: Geografia e História Econômica

Classificação JEL: O13; O47; R11.

Fortaleza, 2020.

* Doutoranda em Economia Rural (PPGER/UFC). winniefinancas@gmail.com. Tel. (88 98129-5368).

** Professor Titular do Departamento de Economia Agrícola (DEA/PPGER/UFC). roberiotcampos@gmail.com.

*** Professor Adjunto do Departamento de Economia Agrícola (DEA/PPGER/UFC). vitormiro@gmail.com.

1 INTRODUÇÃO

A agricultura é um dos pilares da economia brasileira, desde a colonização portuguesa até os dias atuais. Em décadas mais recentes, o setor passou por um processo de intensa modernização, com a mecanização de processos, o aprimoramento das práticas agrícolas e a diversificação da produção. Políticas públicas, investimentos públicos e privados e novas tecnologias permitiram a expansão de áreas agrícolas, da produção e da produtividade, que tornou possível ao Brasil, não apenas melhorar o abastecimento para a população e o agronegócio, mas se firmar como um dos grandes exportadores de produtos agrícolas do mundo. Deste modo, analisar o comportamento da agricultura é de suma importância para o País e suas regiões.

Em virtude da grande dimensão territorial, dos diferentes aspectos climáticos, culturais e socioeconômicos dos estados da região Nordeste do Brasil, torna-se eficaz a análise do progresso do seu setor agrícola como um todo e por microrregiões geográficas, como meio de contribuir positivamente para a literatura ainda escassa. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), essa região é uma das cinco maiores do Brasil, possui 1.554.291,744 km² de área territorial total sendo composta por nove estados. Em função de suas diferentes características físicas, ela se fragmenta em 188 microrregiões, assim distribuídas: Alagoas (13), Bahia (32), Ceará (33), Maranhão (21), Paraíba (23), Piauí (15), Pernambuco (19), Rio Grande do Norte (19) e Sergipe (13).

De acordo com os dados das Contas Regionais do IBGE, o Nordeste foi uma das regiões que mais cresceu em termos de PIB (+3,3% a.a.), no período que compreende os anos de 2002 a 2015. Segundo Leão (2019), entre os estados nordestinos, no período antes aludido, os que mais despontaram em termos de crescimento médio do PIB foram Piauí (+4,8% a.a.), impulsionado pelo crescimento da indústria, Maranhão (+4,5% a.a.), dominado pelos setores da Agropecuária e da Indústria, acompanhado pelo desenvolvimento do cultivo de soja, Paraíba (+4,1% a.a.), em razão da forte indústria extrativa e Ceará (+3,5% a.a.), motivado pelo subsetor de produção de eletricidade e de gás.

Considerando estes aspectos, e com o intuito de contribuir para melhor entender a estrutura espacial e a trajetória do setor agrícola das microrregiões nordestinas, a presente pesquisa se propõe a analisar a estrutura espacial da produtividade da agricultura no período 1996 a 2017, mensurada por meio da razão entre quantidade produzida (em toneladas) e área plantada (em hectares). Uma forma interessante para avaliar o desempenho agrícola é analisar a trajetória de produtividade desse setor ao longo do tempo, o padrão espacial entre regiões e municípios, e por meio dessa mensuração indagar se a sua evolução apresenta algum tipo de convergência.

Nesse sentido, tem-se como objetivo verificar a estrutura espacial e a trajetória do setor agrícola das microrregiões nordestinas, por meio da produtividade agrícola no período 1996 a 2017, além de testar a hipótese de convergência desta. Com base nos resultados, podem-se fornecer informações que subsidiem a formulações de políticas públicas mais efetivas orientadas para o setor, considerando os mecanismos que possam dinamizar a agricultura e que se constituam em estratégias sustentáveis para aperfeiçoá-la.

2 REFERENCIAL TEÓRICO

2.1 A Economia Nordestina

O Nordeste brasileiro é composto por nove estados: Alagoas (AL), Bahia (BA), Ceará (CE), Maranhão (MA), Paraíba (PB), Pernambuco (PE), Piauí (PI), Rio Grande do Norte (RN) e Sergipe (SE), e seu território compreende 1.793 municípios, 42 mesorregiões e 188 microrregiões em uma área de 1.554,3 mil km², correspondentes a 18,3% do território brasileiro, segundo o IBGE. Sua heterogeneidade deriva da imersão de diversas culturas que ao longo dos anos foram se difundindo.

Araújo (1997) destaca como possível agregar o território nordestino em subconjuntos socioeconômicos derivados do processo de ocupação humana e econômica: do Rio Grande do Norte até Alagoas, onde a economia açucareira e a pecuária geravam poderosas oligarquias; do Ceará que era destaque pelo complexo gado-algodão-agricultura de subsistência; do Sergipe à Bahia em que a cana, o cacau e as zonas de combinações agrícolas eram destaques; e do Piauí ao Maranhão que até o final dos anos 1950 era visto como área aberta à expansão da fronteira agrícola regional (SUDENE, 1967).

Historicamente, rotulado como “problemático”, mudanças importantes nas últimas décadas modificaram a realidade da região nordestina. A atual e crescente complexidade da realidade econômica regional permite desvendar uma das mais marcantes características do Nordeste atual: a grande diversidade e a crescente heterogeneidade de suas estruturas econômicas (ARAÚJO, 1997).

O autor dá destaque e atribui como principais fatores as modificações nas últimas décadas da região, como o surgimento do complexo petroquímico de Camaçari (BA), o polo têxtil de Fortaleza (CE), o complexo mineiro-metalúrgico de Carajás (que apesar de se localizar no estado do Pará possui grande influência sobre áreas do Maranhão), o polo agroindustrial de Petrolina-Juazeiro (BA), as áreas de moderna agricultura e grãos do cerrado baiano até o sul de Maranhão e Piauí, o moderno polo de fruticultura com agricultura irrigada no Vale do Açu (RN) e os diversos polos turísticos que se estendem pelo litoral nordestino.

A partir do início do século XIX, com a crise do algodão, proveniente da proliferação do bicudo e das alterações de demanda exterior do produto, a vida da população do semiárido nordestino tornou-se ainda mais difícil, no qual o algodão era representativamente a maior fonte de renda dos pequenos produtores. Diante disso, os produtores se obrigaram a levar ao mercado o escasso excedente da agricultura de sequeiro, em sua maioria, produtora de milho, feijão e mandioca. Araújo (1997) acrescenta que a estratégia brasileira das últimas décadas foi concentrar a expansão da agropecuária em áreas novas (especialmente no Centro-Oeste) e que no Nordeste se assistiu também o grande dinamismo agropecuário e agroindustrial do oeste baiano, do sul do Maranhão e do Piauí.

Os novos polos agrícolas como o oeste baiano, sul do Maranhão e do Piauí, destinados principalmente à produção de soja para atender à demanda externa, têm estabelecidos importantes relações extras regionais, tal qual a produção agroindustrial associada à irrigação no Vale do São Francisco (BA e PE) e no Vale do Açu (RN).

Lima (2012), ao utilizar os dados do Instituto Brasileiro de Geografia e Estatística (IBGE) e do Banco do Nordeste do Brasil (BNB), relata que em 2009 o Produto Interno Bruto (PIB) do Nordeste representava 14,4% do PIB do Brasil, concentrando-se em Bahia (29,4%), Pernambuco (16,8%) e Ceará (14,1%), sendo responsáveis, juntos, por 60,3% do total. Já o Valor Adicionado Bruto (VAB) da Agropecuária ou PIB do setor agropecuário, no mesmo período e região representava 21,1% de toda a produção agropecuária nacional, com destaque

para a Bahia (28,2%) e Maranhão (18%), cujos resultados devem-se à consolidação da área do Cerrado como a grande fronteira agrícola da Região.

Os dados mais recentes do IBGE mostram que em 2017 o VAB do Nordeste representou 18,29% da produção agropecuária nacional. A Tabela 1 mostra o Valor Adicionado Bruto da Agropecuária nordestina, em 2017, por estado. Percebe-se que, assim como em 2009, a Bahia se destaca com o maior VAB da região, seguida por Alagoas.

Tabela 1: Valor Adicionado Bruto (VAB) da agropecuária, por Unidade Geográfica, 2017

Unidade Geográfica	VAB Agropecuária (R\$ mil)	(%)
Nordeste	55.427.281	100
Maranhão	7.508.658	13,55
Piauí	3.791.625	6,84
Ceará	7.487.618	13,51
Rio Grande do Norte	2.160.568	3,90
Paraíba	2.156.214	3,89
Pernambuco	6.036.386	10,89
Alagoas	8.493.021	15,32
Sergipe	1.966.337	3,55
Bahia	15.826.853	28,55

Fonte: IBGE - Produção Agrícola Municipal (2017).

2.2 Fatores importantes para a análise de produtividade agrícola

A importância de analisar a produtividade da agricultura e as mudanças que esta vem passando, justifica-se por ser um aspecto essencial para se entender o crescimento de longo prazo da agricultura (GASQUES *et al.*, 2004). Com base nos Censos Agropecuários de 1970 a 2006, os autores atribuíram o aumento da produtividade agrícola brasileira a alguns pilares como a qualificação da mão de obra ocupada na agricultura, o aumento do número de máquinas e equipamentos, o aumento dos gastos em pesquisa, o crescimento dos subsídios ao crédito rural e a adoção da tecnologia em escala cada vez maior. Evidenciaram que houve crescimento de 243% do produto da agricultura em contraponto ao avanço de 53% no uso de insumos, mostrando que o crescimento da agricultura brasileira foi resultado da evolução da produtividade.

Gasques *et al.* (2004), ao analisarem os condicionantes da produtividade da agropecuária brasileira, no período de 1975 a 2002, utilizaram os fatores terra, pessoal ocupado, máquinas agrícolas automotrizes e uso de fertilizantes e defensivos agrícolas para estimar um índice agregado de insumos. Os resultados apontaram que o crescimento anual médio do índice de produto está associado a acréscimos da Produtividade Total dos Fatores (PTF), em que a terra é pontada como principal responsável para tal. Para os autores, o desempenho da PTF, que mede a relação entre o produto total e o insumo total, é reflexo das elevadas taxas de crescimento da produtividade de mão-de-obra, terra e capital.

Outro ponto importante dos resultados de Gasques *et al.* (2004) diz respeito aos impactos positivos causados com gastos em pesquisa e crédito rural sobre a PTF, em que a variação de 1% nos gastos e em crédito rural causou impactos imediatos da ordem de 0,17% e 0,06% na Produtividade Total dos Fatores, respectivamente. No entendimento de Rezende (2006), com o subsídio do capital, a tecnologia tomou o poder de intensificar a mecanização e facilitar a ocupação de novas áreas como o Cerrado, enquanto a legislação trabalhista e agrária elevou os custos com mão de obra. Desta forma, observou que a produtividade agrícola cresceu cada vez mais à luz do capital e da terra e cada vez menos pela ocupação de mão de obra.

Gasques *et al.* (2018) mostram que o produto agropecuário brasileiro quadruplicou no período de 1975 a 2016 e que esse relevante aumento foi resultante de muitas transformações

ocorridas durante as últimas quatro décadas, tais como a abertura econômica brasileira, na década de 1980, a criação da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), maior utilização de fertilizantes, que têm amenizado as perdas de produto, mudanças na forma de o governo atuar no crédito rural e outras políticas públicas voltadas para o setor.

No tocante ao uso de insumos, Gasques *et al.* (2018) relatam que no período de 1970 a 1980 a terra foi o principal fator de crescimento do produto agropecuário e que, a partir dos anos 1980, até a atualidade, o capital passou a ser o principal fator de crescimento da agricultura brasileira.

Para Bragagnolo e Barros (2015), aumentos da Produtividade Total dos Fatores causaram leve redução do trabalho agrícola, isto porque as inovações tecnológicas empregadas na agricultura por meio de instrumentos como máquinas e tratores, colaboraram para o não crescimento do quantitativo de mão de obra na agricultura brasileira. Por sua vez, as elevações na PTF causaram aumento na área plantada e no emprego de capital. Logo, sugeriram que a modernização da agricultura nas últimas décadas exerceu influência na fronteira agrícola e permitiu a exploração de novas áreas (BRAGAGNOLO; BARROS, 2015).

2.3 Análises Espaciais e de Convergência da Agropecuária

De acordo com Anselin (1996), a econometria convencional não apresenta ferramentas teóricas suficientes para explicar problemas que ocorrem no espaço, como os efeitos espaciais de especificação, na estimação e no teste de hipótese e previsão de modelos que utilizam dados cross-section ou com painel de dados. Para superar tal limitação, a econometria espacial lida com os efeitos provocados pela autocorrelação espacial e, também, pela heterogeneidade espacial, ou estrutura espacial para esses tipos de dados.

A existência ou não desses efeitos pode ser detectada por meio da Análise Exploratória de Dados Espaciais (AEDE), fundamentada no Índice de Moran que identifica a coincidência de valores similares em regiões vizinhas em termos globais, e o indicador local de associação espacial (local indicator of spatial association – Lisa), que detecta a dependência espacial em nível local (ANSELIN, 1996).

Portanto, a AEDE mostra-se aderente à tomada de decisão quanto à alocação de recursos nas atividades produtivas e, ainda, fornece suporte analítico para construção de políticas, planos e projetos de desenvolvimento para regiões. Associada a isso, está a relevância em caracterizar a trajetória comportamental dessas atividades e detectar características e/ou variáveis que as fazem entrar em um processo de convergência.

Solow (1956) iniciou seus estudos sobre convergência propondo um modelo que indicava se as taxas de crescimento da renda per capita de diferentes regiões/países tendiam ao equilíbrio em longo prazo, denominado pelo autor de estado estacionário e, em se confirmando a hipótese, as regiões mais pobres deveriam crescer em taxas maiores que as mais ricas.

Nesse contexto, Lopes (2004) afirma que a convergência é um processo em que uma mesma variável apresenta diferentes valores para diferentes regiões (municípios, estados, países) e que, ao longo dos anos, essa diferença se reduz por motivos variados, como a difusão de tecnologia e a retirada de obstáculos ao crescimento da produção, destacados pela autora. Estudos mostram, por exemplo, que a convergência de renda pode ser entendida como a tendência para uma redução das diferenças de rendimentos em nível mundial (FAGERBERG; GODINHO, 2003).

Barro e Sala-i-Martin (1990), seguindo uma função de produção que incorpora inovação tecnológica, a partir do modelo de Solow (1956), desenvolveram uma expressão algébrica para medir vários tipos de convergência (α , β), procurando destacar diferenças e

similaridades, que até então haviam sido ilustradas apenas graficamente. Segundo os autores, a β -convergência caracteriza uma correlação negativa entre crescimento da renda per capita e o logaritmo da renda per capita inicial, indicando convergência absoluta e/ou condicional entre as economias. Já a α -convergência analisa a redução da dispersão da renda per capita das regiões no período de estudo.

Baumol (1986) empregou a Equação (1), a seguir, para analisar um suposto processo de convergência de renda para 16 países industrializados no período de 1870 a 1979:

$$\frac{1}{T} \left[\ln \left(\frac{y}{n} \right)_{i,t} - \ln \left(\frac{y}{n} \right)_{i,t-1} \right] = \alpha + \beta \ln \left(\frac{y}{n} \right)_{i,t-1} + \varepsilon_i \quad (1)$$

Em que t é o tempo, $\ln(y/n)$ é o logaritmo da renda per capita, ε é o termo de erro e i é o termo indexador para os países. Se existir convergência absoluta, o β deverá ser estatisticamente significativo e apresentar sinal negativo. A convergência de renda tem implicações relevantes sobre o processo de crescimento e desenvolvimento de diferentes regiões e a relação econômica entre elas, bem como sobre a atuação das políticas públicas nesse processo (FREITAS; ALMEIDA, 2015).

Os estudos sobre convergência passaram a incorporar o método da análise espacial com o objetivo de se tornarem cada vez mais próximos da realidade. De acordo com Magalhães *et al.* (2005), a análise econômica está cada vez mais focada em questões relacionadas à dimensão espacial dos problemas, sendo uma característica importante quando a atenção é direcionada à convergência da produtividade regional.

Almeida, Perobelli e Ferreira (2008) ao investigarem sobre a convergência espacial da produtividade agrícola nas microrregiões brasileiras, entre os anos de 1991 e 2003, utilizaram a análise econométrica com a incorporação dos chamados efeitos espaciais, inferiram sobre a importância da agricultura para o crescimento econômico, que tende a ser desigual para as diferentes regiões do país, e chegaram à conclusão de que existem evidências em favor da convergência da produtividade.

Ao analisar dados de produção agrícola, Almeida (2012) concluiu que a autocorrelação ou dependência espacial pode ser ocasionada pelo processo de difusão de práticas/tecnologias agrícolas, cujos efeitos podem ser notados nos locais vizinhos de uma unidade espacial analisada. Em estudos que tratam da convergência da produtividade agrícola, é relevante conhecer a dependência espacial, ou seja, de que maneira e em que nível um local é dependente dos resultados de locais vizinhos.

Com o objetivo de analisar a evolução da produtividade da agropecuária nas microrregiões do Sul do Brasil, Raiher *et al.* (2016) confirmaram a hipótese segundo a qual a distribuição espacial da produtividade da terra não é aleatória e, portanto, estimaram um modelo de convergência utilizando técnicas da econometria espacial. Os resultados obtidos apontaram para hipótese de convergência absoluta e condicional da produtividade da agropecuária.

Diante dos estudos expostos realizados, o trabalho que se desenha busca trazer contribuições para o debate sobre a convergência da produtividade agrícola no Nordeste brasileiro. A relevância prende-se ao fato de, principalmente, se poder focar a análise desta variável para cada microrregião nordestina, até então não realizada, e em razão da técnica de econometria espacial possibilitar incluir na análise as características espaciais das microrregiões.

4 METODOLOGIA

A análise empírica realizada neste trabalho adota um conjunto de métodos de análise estatística espacial, com o objetivo de compreender os padrões de associação entre variáveis estudadas. Inicialmente, aplica-se o conjunto de técnicas estatísticas conhecido como Análise Exploratória de Dados Espaciais (AEDE) para identificar padrões espaciais tais como aglomerações/clusters. Após a aplicação da AEDE, o presente estudo propõe-se analisar as convergências absoluta e condicional das produtividades agrícolas médias, controlando-se os seus efeitos espaciais. O objetivo é verificar a evolução da produtividade agrícola em perspectiva regional.

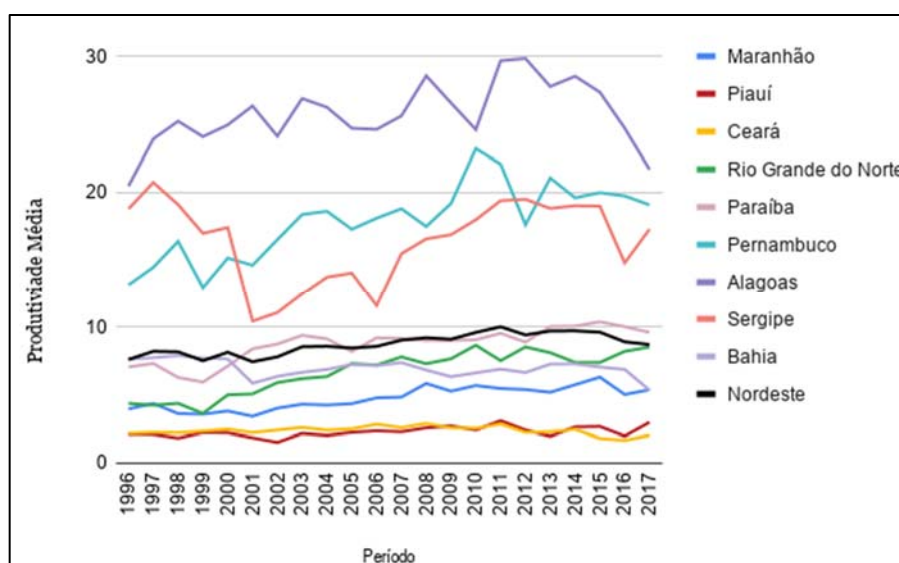
4.1 Base de dados

Diante do objetivo de se analisar espacialmente a produtividade agrícola das microrregiões geográficas do Nordeste brasileiro, faz-se necessário a utilização da técnica chamada de econometria espacial. Para a aplicação deste tipo de análise, é imprescindível o levantamento de dados para todas as microrregiões, pois a ausência dos mesmos pode ocasionar graves problemas. Por conseguinte, o primeiro passo foi verificar a disponibilidade das variáveis: produção agrícola total e a área plantada total de todas as microrregiões durante todo o período a ser observado (1996 a 2017). Após constatar a disponibilidade da Produção Agrícola Municipal (PAM), por meio do Sistema IBGE de Recuperação Automática (SIDRA), a variável produtividade foi criada por intermédio da seguinte razão:

$$P = \frac{\text{produção agrícola total}}{\text{área plantada total}} \quad (2)$$

A Figura 1 mostra a evolução nas últimas duas décadas da produtividade agrícola média das microrregiões geográficas nordestinas e da Região como um todo. É possível notar que apenas três estados se mantiveram bastante acima da média nordestina, sendo eles Alagoas, Pernambuco e Sergipe. Os estados da Bahia, Paraíba e Rio Grande do Norte são os estados que mais se assemelham à trajetória média da Região. Os que obtiveram resultados mais abaixo da média foram Maranhão, Ceará e Piauí. Já a Tabela 2 traz algumas medidas de dispersão da produtividade agrícola de 2017 com respeito às microrregiões por Estado.

Figura 1: Gráfico da produtividade agrícola média dos estados nordestinos (1996-2017)



Fonte: Elaboração própria segundo a Pesquisa Agrícola Municipal (PAM).

Tabela 2: Estatística descritiva da produtividade agrícola média das microrregiões por estado

Unidade Geográfica	Média	Máximo	Mínimo	Desvio Padrão
Nordeste	8,71	51,98	0,13	13,57
Maranhão	5,38	25,46	2,71	4,78
Piauí	2,99	17,81	0,14	4,25
Ceará	2,04	9,80	0,25	2,00
Rio Grande do Norte	8,53	51,98	0,34	12,50
Paraíba	9,61	46,67	0,13	16,74
Pernambuco	19,08	48,65	0,36	20,19
Alagoas	21,68	50,67	0,27	21,73
Sergipe	17,29	40,07	4,64	14,00
Bahia	5,35	39,08	0,48	7,61

Fonte: Elaboração própria segundo a Pesquisa Agrícola Municipal (PAM, 2017).

O Maranhão apresentou baixo desvio padrão, pois apenas uma microrregião, Coelho Neto, obteve produtividade (25,46) muito superior à média, devido à grande produção de cana de açúcar e de soja. O mesmo acontece com Piauí, apenas a microrregião de Teresina (17,81) ficou muito acima da média, devido à grande produção de cana de açúcar. E também no Ceará, que mostrou a menor produtividade média e o menor desvio padrão, apenas a região de Ibiapaba (9,80) se destacou, principalmente pela produção de cana de açúcar, tomate e maracujá.

No Rio Grande do Norte, quatro regiões mostraram produtividades muito acima da média: Litoral Sul (51,98), pela altíssima produção de cana de açúcar; Litoral Nordeste (22,85), pela grande produção de cana de açúcar e de banana; Macaíba (20,79) por produzir também muita cana de açúcar; e Vale do Açu (15,68) por produzir muita banana. Assim como na Paraíba, em que também quatro microrregiões obtiveram as maiores produtividades: Litoral Sul (46,67), Sapé (44,77), João Pessoa (44,06) e Litoral Norte (42,99), resultado da enorme produção de cana de açúcar em ambos e da grande produção de abacaxi na última.

Alagoas, Pernambuco e Sergipe foram os estados com maiores produtividades agrícolas resultantes do maior número de microrregiões com produções muito acima da média graças à produção de cana de açúcar. Em Alagoas os destaques foram: Maceió (50,67), Mata Alagoana (45,84), Serrana dos Quilombos (45,13), Penedo (43,82), São Miguel dos Campos (43,78) e Litoral Norte Alagoano (33,16). Por sua vez, em Pernambuco, os maiores produtores foram: Mata Setentrional Pernambucana (48,65), Recife (47,11), Itamaracá (45,71), Suape (44,94), Vitória de Santo Antão (43,40) e Mata Meridional Pernambucana (39,11). Já em Sergipe foram: Baixo Cotinguiba (40,07), Aracaju (40,01), Cotinguiba (39,25) e Japarutuba (26,37).

A Bahia, apontado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) como o estado nordestino com maior contribuição sob o Valor Adicionado Bruto (VAB) da Agropecuária nordestina, mostrou a soberania de duas microrregiões produtoras: Juazeiro (39,07) resultante da grande produção de cana de açúcar e de manga; e Porto Seguro (23,94) devido à enorme produção de cana de açúcar e mamão.

Tomando-se como base a seção 2.3, em que se discutem os fatores importantes para a análise de produtividade agrícola, e considerando a escassez de informações atualizadas para a aplicação do modelo de convergência condicional, definiram-se algumas variáveis de controle (Tabela 3) como representantes dos fatores de produção a partir dos censos agropecuários de 1996, 2006 e 2017. Assume-se que tais fatores possam representar capital humano e capital físico no modelo, tomando-se por base a Teoria de Crescimento Endógeno.

Tabela 3: Variáveis utilizadas no modelo de convergência condicional

	Identificador	Definição
Variável Dependente		
Produtividade agrícola	P	$\frac{\text{quantidade produzida (t)}}{\text{área plantada (ha)}}$
Variáveis Explicativas		
Assistência técnica	Assistec	$\frac{\text{n}^\circ \text{ de estabelecimentos com assistência técnica regular}}{\text{área explorada (ha)}}$
Área explorada	Area	área explorada (ha)
Mão de obra por	MDO	$\frac{\text{n}^\circ \text{ de pessoas ocupadas}}{\text{área explora (ha)}}$
Número de tratores por hectares	Tratores	$\frac{\text{n}^\circ \text{ de tratores}}{\text{área explorada (ha)}}$

Fonte: Elaboração própria.

Em razão da escassez de informações complementares que pudessem explicar a trajetória de convergência condicional, os períodos verificados foram menores (3) e mais longos (1996-2006, 2006-2017 e 1996-2017), pois os censos agropecuários são realizados a cada 10 anos no País. Para a verificação de convergência absoluta, optou-se em aumentar o número de subperíodos observados (1996-2002, 2003-2007, 2008-2012, 2013-2017 e 1996-2017) contemplando menos observações em cada período, para se averiguar com mais especificidade o comportamento de sua trajetória, pois se torna possível pela disponibilidade anual dos dados por meio da Pesquisa Agropecuária Municipal (PAM).

4.2 Especificação do Modelo Espacial

4.2.1 Matrizes de Ponderação Espacial

Inicialmente, define-se uma matriz de ponderação espacial (W), quadrada de dimensão n por n , para refletir as interações do arranjo espacial do fenômeno a ser estudado. Cada conexão entre duas regiões é representada por uma célula dessa matriz, sendo denominada de peso espacial. Os pesos espaciais w_{ij} representam o grau de conexão entre as regiões segundo algum critério de proximidade, mostrando a influência da região j sobre a região i . Esse grau de conexão pode ser classificado de acordo com o grau de vizinhança baseada na contiguidade, de distância geográfica ou se apoia em critérios socioeconômicos¹.

Neste trabalho, as matrizes testadas baseiam-se nos critérios de contiguidade, em que a matriz de pesos espaciais binários pode ser construída conforme a ideia de que duas regiões são vizinhas, caso elas partilhem de uma fronteira física comum, partindo-se do pressuposto que duas regiões contíguas possuem uma maior interação espacial (ALMEIDA, 2012), pois a ideia que se tem, segundo a Primeira Lei da Geografia, é que lugares mais próximos tenham uma maior interação espacial. Dessa forma, para ambos os critérios, atribui-se valor unitário na matriz a duas regiões vizinhas e valor nulo caso contrário. Formalmente:

$$w_{ij} = \{1 \text{ se } i \text{ e } j \text{ são contíguos}; 0 \text{ se } i \text{ e } j \text{ não são contíguos}\}$$

O problema está em como se define o conceito de fronteira geográfica por intermédio da observação de um mapa, que contém erros de medida por representar abstratamente a real configuração geográfica. Levando-se em conta esses erros e tomando como referência os movimentos de peças num tabuleiro de xadrez, a contiguidade é dita como rainha (do inglês, *Queen*) caso além das fronteiras com extensões diferentes de zero forem considerados os

¹ Nas matrizes socioeconômicas, por exemplo, é possível expressar a similaridade/dissimilaridade entre regiões pela diferença da composição demográfica, diferenças de indicadores sociais ou econômicos (ALMEIDA, 2012).

vértices como contíguos na visualização de um mapa. E será considerada como torre (do inglês, *Rook*) caso apenas as fronteiras físicas com extensões diferentes de zero entre as regiões sejam levadas em conta.

4.2.2 Índice de Moran

A estatística I de Moran é utilizada para identificar a existência de padrões espaciais de interação. Tal ferramenta testará a autocorrelação espacial global entre observações que têm grau de vizinhança especificado pela matriz de pesos espaciais (W). O índice é formalmente definido como:

$$I = \frac{n}{W} \left[\frac{\sum_i \sum_j^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2} \right] \text{ para } i \neq j \quad (3)$$

Em que: n é o número de observações, W é a soma dos ponderadores da matriz, w_{ij} é o elemento na matriz de vizinhança para o par i e j , z_i e z_j são desvios em relação à média ($z_i - z$), ($z_j - z$) e z é a média.

O Índice de Moran fornece três informações: a significância estatística informa se os dados estão distribuídos aleatoriamente ou não; o sinal positivo da estatística Índice de Moran, desde que significativo, indica autocorrelação espacial positiva, revelando similaridade entre os valores do atributo estudado e de sua localização espacial, ou seja, a autocorrelação positiva mostra que altos valores da variável em análise tendem a estar rodeados por altos valores desta em regiões vizinhas.

O sinal negativo, por sua vez significativo, também indica a concentração nas regiões, mas com valores dissimilares, ou seja, que altos valores da variável tendem a ser rodeados por baixos valores e vice-versa. A magnitude da estatística mostra a força de autocorrelação espacial, pois quanto mais próximo de 1 ou -1 mais forte será a autocorrelação e quanto mais próximo de zero, mais dispersos estarão os dados (ALMEIDA, 2012). Ou seja, a interpretação é similar à do teste de correlação de Pearson: 0 indica inexistência de autocorrelação espacial, 1 indica autocorrelação positiva perfeita e -1 autocorrelação negativa perfeita.

O Índice será calculado para identificar a existência de dependência espacial significativa nas variáveis dependentes das regressões. Serão utilizadas as matrizes de contiguidade tipo Queen e Rook de primeira ordem.

4.2.3 Autocorrelação Espacial Local

Do ponto de vista estatístico, a análise global da dependência espacial pode distorcer os resultados em nível local, pois embora o comportamento geral da variável analisada possa ser relevante sob uma perspectiva estatística, esta esconde algumas particularidades presentes em determinadas localizações do conjunto geográfico considerado.

Desta forma, as análises territoriais normalmente estão mais direcionadas para identificar o comportamento local, juntamente com identificação das características próprias de cada espaço analisado. Assim, o método estatístico Indicador Local de Associação Espacial (LISA), proposto por Anselin (1996), será utilizado para verificar a autocorrelação espacial local, o qual utiliza o Índice de Moran para identificar padrões locais de autocorrelação espaciais estatisticamente significativos.

O mapa de LISA, denominado por Anselin (1996) de “*LISA MAP*”, univariado faz uma decomposição do indicador global de autocorrelação na contribuição local de cada observação, em que, basicamente, identifica quatro padrões de agrupamentos, são eles:

- a) Alto-Alto (AA): alto nível de produtividade que se localiza perto de alto nível de produtividade;

- b) Alto-Baixo (AB): alto nível de produtividade que se localiza perto de baixo nível de produtividade;
- c) Baixo-Baixo (BB): baixo nível de produtividade que se localiza perto de baixo nível de produtividade;
- d) Baixo-Alto (BA): baixo nível de produtividade que se localiza perto de alto nível de produtividade.

O coeficiente I_i de Moran local observado para a região i , pode ser expresso pela equação:

$$I_i = Z_i \sum_{j=1}^j W_{ij} Z_j, \text{ para } i \neq j \quad (4)$$

Em que: I_i abrange os vizinhos da observação i , definidos conforme uma matriz de pesos espaciais. Assim, para que I_i seja um indicador LISA é necessário que o somatório dos indicadores locais seja equivalente ao indicador global correspondente, de acordo com um fator de proporcionalidade (ANSELIN, 1996).

Por sua vez, o Índice de Moran local bivariado que investiga a existência ou não de autocorrelação espacial em um contexto bivariado, por exemplo, pode ser interessante para analisar se uma elevada taxa de crescimento da produtividade agrícola de certa região está associada a baixos níveis de produtividade inicial ou a altos níveis de capital físico.

A interpretação intuitiva do Índice de Moran bivariado é a seguinte: i) um valor positivo de I sugere que altos valores da variável z_1 estão associados a altos valores da variável z_2 nas regiões vizinhas; e, ii) um valor negativo de I indica que altos valores da variável z_1 estão associados a baixos valores de z_2 nas regiões vizinhas. Para isso, pode-se calcular a estatística I de Moran para duas variáveis diferentes padronizadas (z_1 e z_2), da seguinte forma:

$$I = \frac{n}{S_0} \frac{z_1 z_2}{z_1' z_1} \quad (5)$$

4.2.4 Convergência Espacial

A hipótese de convergência possui um significado relevante para a teoria do crescimento e desenvolvimento econômico. A convergência de uma variável (como a renda, produtividade ou algum outro indicador de interesse) caracteriza um processo segundo o qual, inicialmente, valores desta variável são desiguais entre países, estados ou regiões, mas essa desigualdade se reduz ao longo do tempo. A redução da desigualdade se dará pelo aumento dos níveis da variável nas localidades ou regiões menos desenvolvidas em taxas superiores às das localidades ou regiões mais desenvolvidas.

Em estudos sobre produtividade agrícola, uma questão natural a ser tratada é a possibilidade das microrregiões com menores produtividades alcançarem os níveis das microrregiões mais produtivas. Para testar tal hipótese, serão utilizados os conceitos de β -Convergência, que se dividem em Absoluta e Condicional.

4.2.4.1 Convergência Absoluta

Os modelos de crescimento neoclássicos assumem a presença de progresso tecnológico exógeno e constante, bem como retornos marginais decrescentes. Sob essas perspectivas, a economia tenderia para um único estado estacionário e as mudanças tecnológicas seriam determinantes do crescimento em longo prazo. Portanto, economias com

produtividades iniciais mais baixas apresentariam uma taxa de crescimento mais alta, segundo os conceitos de convergência absoluta.

Com a finalidade de captar questões inerentes a efeitos de transbordamento entre as regiões, devem-se considerar explicitamente componentes espaciais em sua forma funcional. Portanto, os modelos tradicionais de regressão linear, que não levam em consideração os efeitos espaciais (autocorrelação e heterogeneidade espaciais), não devem ser estimados por intermédio do método dos Mínimos Quadrados Ordinários (MQO), pois as estimativas serão inconsistentes e/ou ineficientes. Assim sendo, ao se estimar o modelo de β -convergência por MQO, busca-se apenas identificar, por meio do teste dos Multiplicadores de Lagrange, qual é a melhor maneira de se estimar a Equação, dada como:

$$\ln\left(\frac{P_t}{P_{t-n}}\right) = \alpha + \beta \ln(P_{t-n}) + u_i \quad (6)$$

Em que:

$\ln\left(\frac{P_t}{P_{t-n}}\right)$ é o logaritmo natural da razão entre a produtividade entre dois anos em análise (por exemplo, produtividade em 2017 e produtividade média em 2003); $\beta \ln(P_{t-n})$ é o logaritmo natural da produtividade média no período inicial; e, u_i é o termo de erro.

Segundo Barro e Sala-i-Martin (1990), para que haja convergência de renda entre as economias é necessário que o β seja negativo, ou seja, é necessário que haja uma correlação negativa entre a renda *per capita* inicial de uma região com a sua taxa de crescimento.

4.2.4.2 Convergência Condicional

Segundo o critério da convergência condicional, as convergências dos padrões de vida e do progresso tecnológico não seriam inevitáveis, mas seriam condicionais aos atributos de cada economia. De acordo com este conceito, as regiões mais produtivas poderiam continuar crescendo a taxas mais elevadas, sugerindo a possibilidade de divergência de produtividade. Entretanto, caso as características das regiões fossem semelhantes, elas convergiriam, em longo prazo, para níveis idênticos.

Assim, as Novas Teorias de Crescimento, como a Teoria de Crescimento Endógeno, por exemplo, analisam a presença ou não de convergência condicional da seguinte forma:

$$\ln\left(\frac{P_t}{P_{t-n}}\right) = \alpha + \beta \ln(P_{t-n}) + X_{t-n}\theta + u_i \quad (7)$$

Em que:

$\ln\left(\frac{P_t}{P_{t-n}}\right)$ representa a variável dependente (taxa de crescimento da produtividade média agrícola); $\ln(P_{t-n})$ a variável explicativa relativa à produtividade média agrícola inicial; X_{t-n} representa a matriz das variáveis de controle (capital humano e capital físico); u_i é o termo de erro aleatório; α representa a constante; β indica o parâmetro que acompanha a variável P_{t-n} ; e, θ é um vetor de coeficientes relativo à matriz de variáveis X_{t-n} .

Nesse modelo, um valor negativo para β indica a existência de convergência condicional, ou seja, a economia, de acordo com as suas características, está caminhando para o seu próprio estado estacionário. Assim, pode-se dizer que as economias convergem em produtividade agrícola desde que possuam parâmetros idênticos e que crescem mais rapidamente quanto maior fosse a sua distância em relação ao seu estado estacionário (GROLLI *et al.*, 2006).

4.2.5 Modelos Espaciais

4.2.5.1 Modelo Autorregressivo Espacial (SAR)

Um dos modelos mais utilizados para modelagem espacial é o Modelo Autorregressivo Espacial (do inglês, *Spatial Autorregressive Model*²²), ou simplesmente modelo SAR. Tal mecanismo tem como ideia principal incluir um termo de defasagem (em inglês, *lag*) nos regressores da equação, assim como é feito nos modelos Autorregressivos (AR) em séries temporais. Na sua forma mais simples, o modelo SAR tem a expressão:

$$y = \rho W y + \varepsilon \quad (8)$$

Em que:

y é um vetor coluna, contendo n observações na amostra para a variável resposta; o coeficiente escalar ρ corresponde ao parâmetro autorregressivo, esse parâmetro interpreta-se como o efeito médio da variável dependente relativamente à vizinhança espacial na região em questão; o termo ε corresponde aos resíduos da equação; e, W representa a matriz de pesos espaciais.

Dentre as sugestões para a estimação do coeficiente ρ , uma delas é utilizar o estimador de Mínimos Quadrados Ordinários (MQO). No entanto, sabe-se que o estimador de MQO é inconsistente ao relacionar o vetor de covariáveis (variáveis do lado direito da equação) com o resíduo da regressão.

O modelo SAR estendido para incorporar variáveis exógenas no lado direito da equação, apresenta a seguinte expressão:

$$y = \rho W y + X\beta + \varepsilon \quad (9)$$

Em que:

ρ é o parâmetro autorregressivo; W a matriz de vizinhança; X são variáveis independentes; e, ε um termo aleatório de erro normalmente distribuído.

Quando ρ é positivo há indicação de autocorrelação positiva; quando ρ é negativo, a interpretação é de que há autocorrelação negativa.

Desse modo, incluindo à Equação (6), na fórmula funcional do modelo SAR, tem-se:

$$\ln\left(\frac{P_t}{P_{t-n}}\right) = \rho W \left[\ln\left(\frac{P_t}{P_{t-n}}\right) \right] + \alpha + \beta \ln(P_{t-n}) + \varepsilon_i \quad (10)$$

4.2.5.2 Modelo de Erros Espaciais (SEM)

Os modelos SEM (do inglês, *Spatial Error Models*) possuem a seguinte especificação:

$$y = X\beta + u \quad (11)$$

Em que os erros da equação observada possuem uma estrutura autorregressiva da forma:

$$u = \lambda W u + \varepsilon \quad (12)$$

O vetor de erros ε possui distribuição normal multivariada e média nula. O coeficiente escalar λ indica a intensidade da autocorrelação espacial entre os resíduos da equação

²²e³ Ver Anselin (1988) e Lesage e Pace (2008).

observada. Mais especificamente, esse parâmetro mensura o efeito médio dos erros dos vizinhos em relação ao erro da região em questão. Note-se que, ao contrário dos modelos SAR, os modelos SEM não apresentam a variável dependente como uma função direta dos seus *lags* espaciais, mas sim autocorrelação espacial nos termos de erro.

Substituindo a Equação (15) na Equação (6), obtém-se a fórmula funcional do modelo SEM para o presente estudo:

$$\ln\left(\frac{P_t}{P_{t-n}}\right) = \alpha + \beta \ln(P_{t-n}) + (I - \lambda W)\varepsilon_i \quad (13)$$

4.2.5.3 Modelo Autorregressivo Espacial e Média Móvel (SARMA)

Os modelos SEM e SAR podem ser combinados em uma especificação mais geral por intermédio do modelo SARMA (do inglês, *Spatial Autorregressive and Moving Average*) que tem a seguinte especificação:

$$y = X\beta + (I - \lambda W)^{-1}\varepsilon \quad (14)$$

Em que, segundo Albuquerque *et al.* (2008), ε possui distribuição normal multivariada com média nula e covariância $\sigma^2 I$. Portanto, o vetor de variável dependente y possui distribuição normal multivariada com média condicional $E[y|X] = X\beta$ e matriz de variância condicional $\Sigma_{y|X} = \sigma^2(I - \lambda W)^{-1}[(I - \lambda W)^{-1}]^T$.

Substituindo a Equação (17) na Equação (6), obtém-se a expressão funcional do modelo SARMA:

$$\ln\left(\frac{P_t}{P_{t-n}}\right) = \alpha + \beta \ln(P_{t-n}) + (I - \lambda W)^{-1}\varepsilon_i \quad (15)$$

5 RESULTADOS

5.1 Análise do Índice de Moran

Segundo o critério de Baumont (2004), as significâncias das matrizes de pesos devem ser analisadas fazendo um paralelo entre as variáveis dependentes em estudo. Desse modo, a Tabela 4 mostra que os melhores resultados estão nas matrizes de contiguidade *Rook* de ordem 1, as quais mostram maiores índices de Moran significativos para todos os períodos em análise. Portanto, os resultados, a partir de agora, foram modelados seguindo a melhor opção mostrada da matriz de ponderação espacial.

Tabela 4 : Índice de Moran da produtividade agrícola total das microrregiões por subperíodos

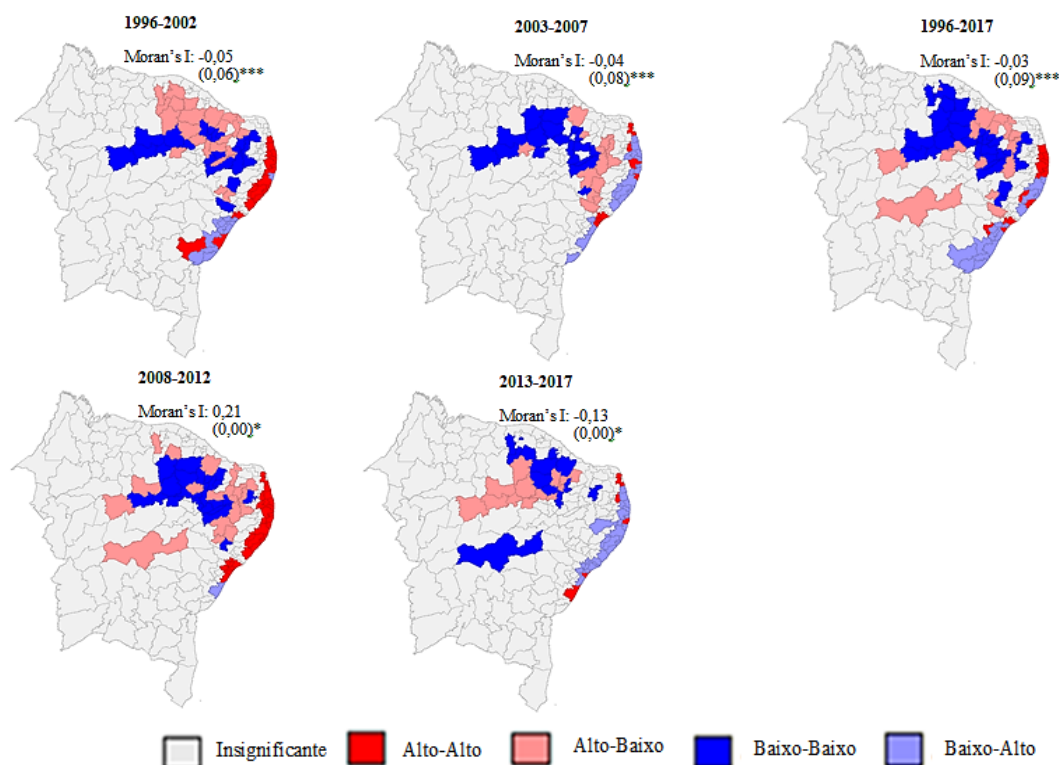
W	Subperíodos para Análise Absoluta					Subperíodos para Análise Condicional		
	1996-2002	2003-2007	2008-2012	2013-2017	1996-2017	1996-2006	2006-2017	1996-2017
Queen	0,295	0,277	0,293	0,279	0,293	0,309	0,320	0,293
p-valor	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Rook	0,297	0,279	0,293	0,282	0,295	0,311	0,321	0,295
p-valor	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Fonte: Elaboração própria.

5.1. Análise de Autocorrelação Espacial

A Figura 2 apresenta os mapas de *clusters* LISA bivariados (BiLISAs) para a taxa de crescimento da produtividade total agrícola e a produtividade agrícola total inicial nos subperíodos da análise de convergência absoluta.

Figura 2 : BiLISAs para a taxa de crescimento da produtividade total agrícola e a produtividade agrícola total inicial das microrregiões do Nordeste por subperíodos



Fonte: Elaboração própria.

Legenda: *= 1% de significância, ***= 10% de significância.

A existência da dependência espacial entre taxa de crescimento da produtividade total agrícola e a produtividade agrícola total inicial, é verificada pela significância estatística em todos subperíodos analisados. O sinal positivo do Índice de Moran (em 2008-2012) revela que a produtividade está concentrada ao longo do território do Nordeste. Já em 1996-2002, 2003-2007, 2013-2017 e 1996-2017, o sinal negativo mostra que a produtividade agrícola não está concentrada ao longo do território, mas possui alguns pontos de aglomeração (*clusters*).

Cabe destacar que a correlação negativa existente entre essas variáveis é um indício de que há um processo de convergência absoluta de produtividade, ou seja, a hipótese é de que, ao longo do período, há uma tendência de redução das desigualdades entre a produtividade das microrregiões nordestinas. Significa ainda que microrregiões com altas produtividades agrícolas no período inicial possuem vizinhos com baixas taxas de crescimento e vice-versa.

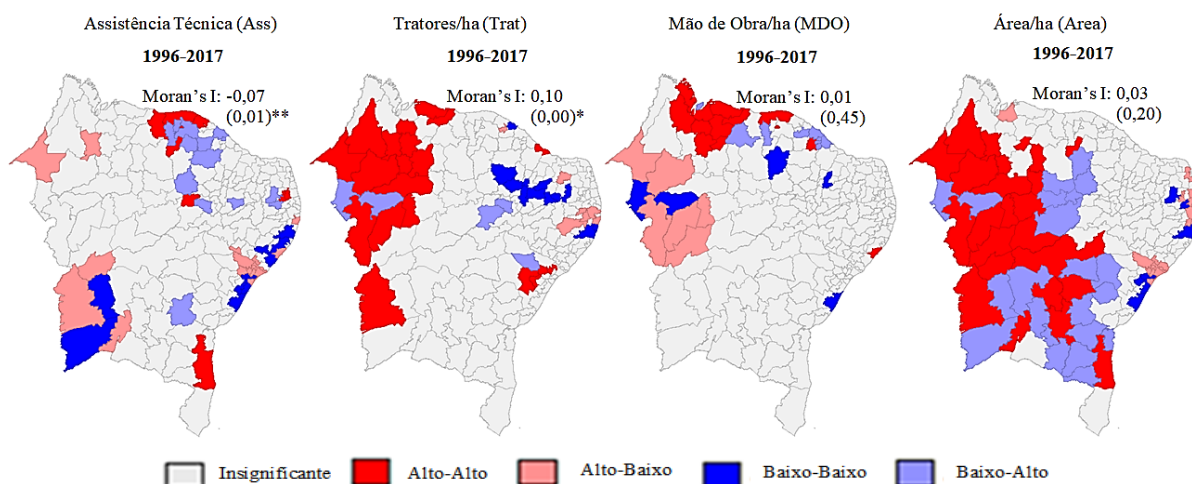
Analisando-se o comportamento dos agrupamentos da última figura, que representa o período completo (1996-2017), percebe-se que a alta produtividade ou Alto-Alto (AA) está concentrada em todos os períodos nas microrregiões litorâneas dos estados de Rio Grande do Norte (Litoral Norte), Paraíba (João Pessoa, Sapé), Pernambuco (Itamaracá, Litoral Sul, Mata Setentrional Pernambucana, Recife, Vitória de Santo Antão), Alagoas (Maceió, Serrana dos Quilombos), Sergipe (Cotinguiba, Aracajú, Japarutuba).

Também foram identificadas várias microrregiões de alta produtividade cercadas por microrregiões de baixa produtividade, como Baixo Jaguaribe (CE), Chapada do Araripe (CE), Várzea Alegre (CE), Médio Jaguaribe (CE), Chapada do Apodi (RN), Mossoró (RN), Vale do Açu (RN), Serra de Santana (RN), Seridó Oriental (RN), Cariri Ocidental (PB), Itaporanga (PB), Floriano (PI), Serrana do Sertão Alagoano (AL), Sergipana do Sertão do São Francisco

(SE) e Juazeiro (BA). A maior parte das áreas detectadas são *clusters* baixa produtividade (Baixo-Baixo ou BB), principalmente nos estado de Ceará, Piauí e Paraíba.

A Figura 3 mostra os LISA MAPs bivariados entre as taxas de crescimento da produtividade total agrícola e das variáveis explicativas para o modelo em que se analisa a convergência condicional.

Figura 3 : BiLISAs para a taxa de crescimento da produtividade total agrícola e as variáveis explicativas do modelo nas microrregiões do Nordeste



Fonte: Elaboração própria.

Legenda: *= 1% de significância, **=5% e ***= 10% de significância.

Em razão da significância do Índice de Moran para Assistência Técnica (Assistec) e Tratores (Tratores), pode-se observar que existe autocorrelação espacial entre a taxa de crescimento da produtividade total agrícola da microrregião com as taxas de crescimento de tais variáveis explicativas das suas vizinhas. O contrário acontece com Mão de Obra (MDO) e Área Explorada (Area), que se mostrou insignificante quando analisada isoladamente, portanto, nada se pode inferir a respeito destas variáveis.

A estatística de Assistência Técnica igual a -0,07 sugere correlação negativa entre as variáveis, ou seja, indicam que microrregiões com alto nível do fator capital físico inicial possuem vizinhos com baixas taxas de crescimento e vice-versa. Isto indica convergência condicional, significando que, quando condicionada a tal fator, a desigualdade na produtividade agrícola cearense tende a diminuir em longo prazo. O mesmo não acontece com tratores em que, quando analisada isoladamente, seu sinal positivo não indica convergência, apesar de significativo.

Observa-se que as microrregiões com agrupamentos de alta produtividade correlacionada à Assistência Técnica são: Litoral Piauiense (PI), Litoral de Camocim e Acaraú (CE), Itapipoca (CE), Meruoca (CE), Ipu (CE), Chapada do Araripe (CE), Barro (CE), Guarabira (PB), Brejo Paraibano (PB) e Ilhéus-Itabuna (BA). Deste modo, por hipótese, essas microrregiões foram as que mais contribuíram em termos de assistência técnica para a diminuição das desigualdades na produção agrícola nordestina.

5.3 Análise dos Testes e Diagnósticos

O que diferencia a econometria espacial da tradicional é a agregação dos efeitos espaciais no modelo. Para Almeida (2012), tal incorporação gera violações de hipóteses do Modelo Clássico de Regressão Linear, como a presença de dependência espacial que viola a hipótese de média condicional zero, que pode acarretar na ineficiência do estimador, pois os erros são dependentes entre pares de regiões e induz a emergência de heterocedasticidade. Em

busca do modelo mais adequado, aplicou-se o Teste dos Multiplicadores de Lagrange na regressão por meio de MQO. Para os subperíodos 1996-2002, 2003-2007, 2013-2017, 1996-2017 (absoluto) e 1996-2006, 2006-2017, 1996-2017 (condicional), o modelo mais adequado é o SEM, e para o subperíodo 2008-2012, o melhor modelo é o SAR. Tais resultados foram corroborados pelos critérios de informação *Akaike Information Criterion* (AIC), *Schwarz Criterion* (SC) e do valor da função de verossimilhança.

O conjunto de testes para averiguar a presença de autocorrelação espacial é útil tanto para auxiliar a identificação do modelo econométrico espacial quanto para a tarefa de validação ou diagnóstico desse modelo, como a existência de heterocedasticidade. Conforme se verifica na aplicação do teste Breusch-Pagan, considerando um nível de significância de 5%, não se rejeita a hipótese de homocedasticidade dos erros. Logo, assume-se a ausência de heterocedasticidade em quase todos os períodos, com exceção apenas de 2008-2012. Os resultados dos modelos são discutidos a seguir.

5.4 Análise das Convergências

5.4.1 Convergência Absoluta

Após seguir todos os critérios brevemente explicitados, foi possível estimar o modelo de β -convergência absoluta, tal como pode ser observado na Tabela 5.

Tabela 5 : Resultados econométricos dos modelos SEM e SAR para convergência absoluta por subperíodo

Coeficientes	Modelo SEM				Modelo SAR
	1996-2002	2003-2007	2013-2017	1996-2017	2008-2012
α	0.11 (0,14)	0.18* (0,00)	0.18*** (0,09)	0.29** (0,02)	-0.40* (0,00)
β	-0.09* (0,00)	-0.10* (0,00)	-0.22* (0,00)	-0.19* (0,00)	0.13* (0,00)
λ	0.57* (0,00)	0.55* (0,00)	0.60* (0,00)	0.57* (0,00)	0.47* (0,00)
Breusch Pagan	0.27 (0,60)	3.84 (0,11)	1.97 (0,16)	2.82 (0,12)	17.83 (0,00)
LIK	-2479.09	-85.35	-167.69	-204.62	-147.91
AC	4962.19	174.716	339.39	413.25	299.82
SC	4973.18	181.189	345.87	419.72	306.28

Fonte: Elaboração própria.

Legenda: *= 1% de significância, **=5% e ***= 10% de significância.

Nos subperíodos 1996-2002, 2003-2007, 2013-2017 e 1996-2017, as estatísticas de β sinalizaram uma relação negativa, apresentando convergência absoluta, segundo Baumol (1986). Ou seja, as microrregiões com produtividade agrícola inicial maior terão menores taxas de crescimento, convergindo para uma maior paridade em longo prazo. O contrário acontece no subperíodo 2008-2012, em que o β é positivo, não indicando convergência, mas, considerando um nível de significância de 5%, não se pode assumir a ausência de heterocedasticidade nesse período. Logo, nada se pode interpretar dele a priori.

Para os parâmetros estatisticamente significativos e que indicaram convergência absoluta, permite-se calcular as taxas de convergências (Tabela 6).

Tabela 6: Taxas de Convergências Absolutas dos subperíodos estatisticamente significativos

Subperíodo	Taxa de Convergência Absoluta*
1996-2002	0,013
2003-2007	0,021
2013-2017	0,049
1996-2017	0,009

Fonte: Elaboração própria.

*A taxa de convergência (θ) foi computada segundo a fórmula: $\theta = \ln(\beta + 1)/(-k)$, em que k é o número de anos do período (REY E MONTOURI, 1999).

Analisando-se o período como um todo (1996-2017), conclui-se que a taxa de convergência absoluta mostrou-se crescente em aproximadamente 1%, significando que em longo prazo, há uma tendência na diminuição da desigualdade entre a produtividade das microrregiões nordestinas e que, a uma taxa de 1%, a produtividade destas converge para uma média ou um estado estacionário. Vale ressaltar que as taxas se mostraram crescentes na sequência temporal dos subperíodos intermediários, que passa de 1,3%, em 1996-2002, para 2,1%, em 2003-2007, e para 4,9%, em 2013-2017.

5.4.2 Convergência Condicional

Uma vez detectada a convergência absoluta, torna-se interessante verificar de que forma os fatores de produção influenciam neste contexto, por meio de variáveis que os representem. Neste sentido, na Tabela 7 constam os resultados obtidos por meio do modelo que leva em consideração a inclusão de variáveis, as quais buscam condicionar o crescimento da taxa de produtividade agrícola.

Tabela 7 : Resultados econométricos do modelo SEM para Convergência Condicional por subperíodo

Coeficientes	Modelo SEM		
	1996-2006	2006-2017	1996-2017
α	0,37* (0,00)	-0,48* (0,00)	-0,16 (0,83)
β	-0,14* (0,00)	0,27* (0,00)	-0,17* (0,00)
λ	0,52* (0,00)	0,57* (0,00)	0,56* (0,00)
Assistec	1,37 (0,52)	-0,02 (0,68)	-0,01 (0,87)
Tratores	1,37* (0,03)	0,03 (0,32)	0,10 (0,20)
MDO	0,11** (0,03)	-0,05 (0,19)	0,19 (0,29)
Area	-0,04 (0,52)	-0,02 (0,28)	0,02 (0,57)
Breusch Pagan	7,74 (0,17)	7,30 (0,20)	0,10 (0,11)
LIK	-101,85	-138,80	-202,35
AC	215,71	289,60	416,70
SC	235,10	308,99	436,09

Fonte: Elaboração própria.

Legenda: * = 1% de significância, ** = 5% e *** = 10% de significância.

Pode-se observar que em todos os períodos, o β estimado é estatisticamente significativo. O coeficiente negativo e significativo (ao nível de 1%) de β , nos subperíodos 1996-2006 e 1996-2017, indica um processo de convergência condicional, ou seja, as microrregiões

estariam convergindo para a sua produtividade agrícola de equilíbrio, dada suas características específicas. Nesse caso, duas microrregiões só alcançariam produtividades idênticas caso apresentassem, no período inicial, características idênticas.

No que se refere às variáveis explicativas, apenas no primeiro subperíodo, Tratores e Mão de Obra foram estatisticamente significantes ao nível de 5%. Os coeficientes positivos destas retratam uma relação análoga com a variável a ser explicada. Ou seja, em longo prazo, à medida que as taxas de tratores e mão de obra crescem, a taxa de produtividade agrícola total também será crescente, o contrário é recíproco.

Para os subperíodos que indicaram β negativo (convergência condicional) e estatisticamente significante, são então calculadas as taxas de convergências, apresentadas na Tabela 8.

Tabela 8 : Taxas de Convergências Condicionais para os subperíodos estatisticamente significativos, Nordeste

Subperíodo	Taxa de Convergência Condicional*
1996-2006	0,014
1996-2017	0,008

Fonte: Elaboração própria.

*A taxa de convergência (θ) foi computada segundo a fórmula: $\theta = \ln(\beta + 1)/(-k)$, em que k é o número de anos do período (REY E MONTOURI, 1999).

Deste modo, quando são agregadas variáveis que simbolizam os fatores de produção (Assistec, Tratores, MDO e Area), que teoricamente influenciam no crescimento da taxa de produtividade agrícola, a taxa de convergência condicional permanece praticamente a mesma do modelo absoluto, aproximadamente 1%. Então, pode-se inferir, *a priori*, que em um espaço muito grande e heterogêneo como a região Nordeste, as variáveis condicionantes levadas aqui em questão influenciam na diminuição das desigualdades da produção agrícola total, mas de maneira sutil.

CONCLUSÃO

A pesquisa aqui apresentada propôs uma análise do comportamento da produtividade agrícola das microrregiões do Nordeste, buscando evidências a respeito do padrão espacial da produtividade e da existência de convergência desta variável.

Cabe destacar que uma das limitações deste estudo se refere ao uso de microrregiões, uma vez que o ideal seria o uso de municípios, dado que o número maior de observações retrata cada vez mais o real efeito da convergência. Porém, atualmente isso é inviável devido à indisponibilidade de dados. Portanto, trabalhos futuros podem enriquecer os resultados aqui encontrados se realizarem a análise por municípios, além da agregação de mais variáveis no modelo de convergência condicional.

A Análise Exploratória dos Dados Espaciais (AEDE) confirmou a hipótese segundo a qual a distribuição espacial da produtividade da terra não é aleatória, ao identificar que microrregiões com elevada produtividade são rodeadas por microrregiões com produtividade também elevada, e microrregiões com produtividade baixa são cercadas por microrregiões de baixa produtividade. E, mediante a análise do mapa de clusters ou LISA, foi possível identificar alguns agrupamentos principalmente nos Estados de Rio Grande do Norte, Paraíba, Pernambuco e Sergipe.

Constatou-se um processo de convergência absoluta da produtividade agrícola em aproximadamente 1%, significando que em longo prazo, a uma taxa de 1%, a produtividade das

microrregiões nordestinas converge para uma média ou um estado estacionário. Além disso, constatou-se um processo de dependência espacial entre as microrregiões, evidenciando que o crescimento da produtividade de uma determinada microrregião afeta, positivamente, o crescimento da produtividade das microrregiões circunvizinhas.

A hipótese de convergência condicional também foi confirmada para as microrregiões do Nordeste, assim como a taxa de aproximadamente 1%. Isso porque as variáveis estruturais inseridas no modelo não mostraram significância suficiente para explicar a convergência condicional da produtividade.

Assim, as evidências apresentadas na análise preliminar a respeito da mensuração da convergência β da produtividade agrícola da região Nordeste, pode contribuir para a literatura regional em dois aspectos relevantes: a) a incorporação dos efeitos espaciais explícitos no modelo, ou seja, mensura a convergência do setor agrícola, levando-se em conta a interação das regiões com o seu entorno, o que permite verificar a existência de polos e agrupamentos; e b) em contraste com a grande maioria dos trabalhos que tratam habitualmente sobre convergência de renda, inova na utilização da produtividade agrícola como variável de convergência.

A evidência de dependência espacial, detalhada nos mapas de LISA, entre as microrregiões implica que os ganhos em termos de políticas não são restritos a uma determinada localidade, mas, sim, são transbordados para as regiões vizinhas. Neste sentido, pode-se orientar a adoção de polos regionais de desenvolvimento agrícola. Este tipo de estratégia, adotando-se um referencial espacial/geográfico e a definição de polos especializados, pode permitir que a alocação de recursos seja mais eficiente.

Dada a grande heterogeneidade do território brasileiro, torna-se imprescindível a produção de conhecimento que oriente estratégias de desenvolvimento em nível regional, incorporando características próprias de cada região. Espera-se que os resultados obtidos possam incentivar ainda mais o desenvolvimento de análises sobre a evolução da produtividade agrícola.

BIBLIOGRAFIA

ALBUQUERQUE, P. H. M. *et al.* **Conglomerados espaciais**. 2008. 117f. Dissertação (Curso de Pós-Graduação em Estatística) — Universidade Federal de Minas Gerais, Minas Gerais, 2008.

ALMEIDA, E. **Econometria espacial aplicada**. Campinas: Alínea, 2012.

ALMEIDA, E. S. de; PEROBELLI, F. S.; FERREIRA, P. G. C. Existe convergência espacial da produtividade agrícola no Brasil? **Revista de Economia e Sociologia Rural**, Brasília, v. 46, n. 1, p. 31 – 52, 2008.

ANSELIN, L. *et al.* Simple diagnostic tests for spatial dependence. **Regional Science and Urban Economics**, v. 26, p. 77 – 104, 1996.

ARAÚJO, T. B. D. Dossiê Nordeste I. **Estudos Avançados**, USP, São Paulo, v.11, n. 29, p. 1 – 32, 1997.

BARRO, R. J.; SALA-I-MARTIN, X. Economic growth and convergence across the United States. **Journal of Political Economy**, v. 86, n. 4, p. 549 – 580, 1990.

BAUMOL, W. J. Productivity growth, convergence, and welfare: what the long-run data show. **American Economic Association**, v. 76, n. 5, p. 1072 – 1085, 1986.

- BAUMONT, C. Spatial effects in housing price models: do housing prices capitalize urban development policies in the agglomeration of Dijon (1999)?. **Research Report** – Laboratoire d'économie et de gestion (LEG). Dijon, 2004. 26 p.
- BRAGAGNOLO, C.; BARROS, G. S. de C. Impactos dinâmicos dos fatores de produção e da produtividade sobre a função de produção agrícola. **Revista de Economia e Sociologia Rural**, Piracicaba-SP, v. 53, n. 1, p. 31 – 50, 2015.
- FAGERBERG, J.; GODINHO, M. Innovation and catching-up. In: J., F.; D.C., M.; R, N. (ed.). **The Oxford Handbook of Innovation**. Ottawa: Oxford University Press, p. 514 – 544, 2003.
- FREITAS, M. V. de; ALMEIDA, E. Existe realmente convergência de renda entre países? **Estudos Econômicos**, São Paulo, v. 45, n. 2, p. 287 – 316, 2015.
- GASQUES, J. G. et al. **Condicionantes da produtividade da agropecuária brasileira**. Brasília: IPEA, 2004. 33 p. (Texto Para Discussão, n.1017).
- GASQUES, J. G.; BACCHI, M. R. P.; BASTOS, E. T. **Crescimento e Produtividade da Agricultura Brasileira de 1975 a 2016**. Brasília: IPEA, 2018, 9 p. (Nota Técnica IV).
- GROLI, P. A, OLIVEIRA, C. A. e JACINTO, P. A. Crescimento econômico e convergência com a utilização de regressões quantílicas: um estudo para os municípios do Rio Grande do Sul (1970 a 2001). *In: XXXIV Encontro Nacional de Economia – ANPEC. Anais*. 2006.
- LEÃO, H. C. R. S. PIB de seis Estados do Nordeste cresceram acima da média do País. **Diário Econômico ETENE**. Fortaleza: BNB, ano II, n. 19, p. 1-2, 2019.
- LESAGE, J.; PACE, R. K. An introduction to spatial econometrics. **Revue d'économie industrielle**, n. 123, p. 19 – 44, 2008.
- LIMA, L. D.; MONTE, K. A. do; MILITÃO, V. B. **Nordeste em mapas 2012**. Fortaleza: Banco do Nordeste, 2012. (Relatório Técnico).
- LOPES, J. L. **Avaliação do processo de convergência da produtividade da terra na agricultura brasileira no período de 1960 a 2001**. 2004. 208f. Tese (Escola Superior de Agricultura Luiz de Queiroz), Universidade de São Paulo, São Paulo, 2004.
- MAGALHÃES, A.; HEWINGS, G.; AZZONI, C. Spatial dependence and regional convergence in Brazil. **Investigaciones Regionales – Journal of Regional Research**, Barcelona, n.6, p. 5 – 20, 2005.
- RAIHER, A. P. et al. Convergência da produtividade agropecuária do sul do Brasil: uma análise espacial. **Revista de Economia e Sociologia Rural**, Brasília, v. 54, n. 3, p. 517 – 536, 2016.
- REZENDE, G. C. de. Políticas trabalhista, fundiária e de crédito agrícola no Brasil: uma avaliação crítica. **Revista de Economia e Sociologia Rural**, Rio de Janeiro, v. 44, n. 1, p. 47 – 78, 2006.
- REY, S. J.; MONTOURI, B. D. US regional income convergence: A spatial econometric perspective. **Regional Studies**: Cambridge Vol. 33, Ed. 2, p. 143-156, 1999.
- SOLOW, R. M. A contribution to the theory of economic growth. **The Quarterly Journal of Economics**, The MIT Press, Cambridge, v. 70, n. 1, p. 65 – 94, 1956.
- SUDENE. **Uma política de desenvolvimento econômico para o Nordeste**. 2. ed. Rio de Janeiro: Departamento de Imprensa Nacional, v. 2, 1967.

DESMATAMENTO DE MATA ATLÂNTICA NO ESTADO DO CEARÁ: ANÁLISES DA CURVA DE KUZNETS AMBIENTAL A PARTIR DE DADOS EM PAINEL, 2011 A 2017

Wesley Leitão de Sousa*

Guilherme Irffi**

Moisés Dias Gomes de Asevedo***

RESUMO

Esta pesquisa tem por objetivo analisar a relação entre desmatamento da Mata Atlântica e atividade econômica, sob a hipótese da Curva de Kuznets Ambiental, considerando os municípios do estado do Ceará. Para isto, estima-se um modelo *tobit* para dados em painel, no período de 2011 a 2017, considerando o *PIB per capita* e controlando por fatores associados tanto ao desmatamento quanto a proteção ambiental. Ao contrário do esperado, a densidade populacional e a bovinocultura suavizam a atividade do desmatamento. As evidências encontradas para a relação entre desmatamento e *PIB per capita* rejeitam a hipótese de “U-invertido”, e sim, sugerem o formato de “N”, indicando que o desmatamento na região pode ser cíclico. Desta forma, não existem indicativos para assegurar que a atividade econômica dos municípios analisados garanta por si só a proteção ambiental e o uso sustentável da Mata Atlântica e ecossistemas associados no estado do Ceará.

Palavras-chaves: Ceará, Curva de Kuznets Ambiental, Desmatamento, Mata Atlântica.

Classificação JEL: C34, Q50, Q56.

DEFORESTATION OF ATLANTIC FOREST IN THE STATE OF CEARÁ: ANALYSIS OF THE ENVIRONMENTAL KUZNETS CURVE FROM PANEL DATA, 2011 TO 2017

ABSTRACT

This research aims to analyze the relationship between the deforestation of the Atlantic Forest and economic activity, under the hypothesis of the Environmental Kuznets Curve, considering the municipalities of the state of Ceará. For this, it is estimated a *tobit* model for panel data, in the period from 2011 until 2017, considering the *GDP per capita* and controlling for associated factors with both deforestation and environmental protection. Unlike the expectations, the population density and cattle farming soften deforestation activity. The evidence found for the relationship between deforestation and *GDP per capita* rejects the “U-inverted” hypothesis, and yes, suggests the “N” format, indicating that deforestation in the region may be cyclical. Therefore, there are no indications to ensure that the economic activity of the municipalities analyzed assure by itself the

* Doutorando em Economia pela Universidade Federal do Ceará (UFC), Avenida da Universidade 2762, Benfica, Fortaleza, Ceará, 60.020-181, Brasil. E-mail: wesleyleitao@alu.ufc.br. Tel: (85) 98884-5470.

** Professor do Departamento de Economia Aplicada e do Programa de Pós-Graduação em Economia (CAEN) na UFC, Avenida da Universidade 2762, Benfica, Fortaleza, Ceará, 60.020-181, Brasil. E-mail: irffi@caen.ufc.br.

*** Mestre em Economia Rural pela Universidade Federal do Ceará (UFC), Avenida Mister Hull 2977, Campus do Pici, Bloco 826, Fortaleza, Ceará, 60.440-970, Brasil. E-mail: moisesdga@gmail.com.

environmental protection and sustainable use of the Atlantic Forest and associated ecosystems in the state of Ceará.

Keywords: Ceará, Environmental Kuznets Curve, Deforestation, Atlantic Forest.

JEL Classification: C34, Q50, Q56.

Área 03: Geografia e História Econômica.

1 INTRODUÇÃO

A degradação dos ecossistemas florestais se dá em razão de diversos fatores, sobretudo das atividades de origem antrópica. Nesse sentido, há uma intensa preocupação, de que, com o avanço do desenvolvimento econômico, a pressão ambiental exercida aos ecossistemas supere a capacidade de suporte do meio ambiente. A relação entre desenvolvimento econômico e degradação ambiental encontra evidência na Curva de Kuznets Ambiental (CKA), proposta inicialmente em Grossman e Krueger (1995).

A premissa da CKA é de que, em baixos níveis de desenvolvimento, o crescimento da renda *per capita* (*pc*) induz a um aumento da degradação ambiental. No entanto, a partir de certo nível de renda, essa lógica se reverteria, com o aumento da renda *pc* levando a uma redução da degradação ambiental, caracterizando uma curva no formato de “U-invertido”. Assim, Grossman e Krueger (1995) relacionaram a renda *pc* aos indicadores de qualidade da água e do ar, e encontraram evidências em diversos países desenvolvidos, de que a degradação ambiental se eleva com o incremento econômico até um ponto de inflexão a partir do qual ocorre uma melhora na qualidade ambiental resultando na CKA.

No Brasil, inúmeras pesquisas desenvolvem uma relação entre desenvolvimento econômico e indicadores de degradação ambiental. Assim, uma das pautas ecológicas mais pertinentes, diz respeito ao desmatamento em biomas nacionais, tema com foco nos domínios da Amazônia e Cerrado (COLUSSO, PARRÉ & ALMEIDA, 2012; OLIVEIRA, ALMEIDA, FREGUCLIA & BARRETO, 2011) especialmente pela extensão territorial destes ecossistemas, da expansão da fronteira agrícola nestes locais e da disponibilidade de informações. Por outro lado, observa-se um *gap* nas pesquisas que relacionam desenvolvimento econômico e alterações na cobertura florestal nos demais biomas nacionais – Caatinga, Mata Atlântica, Pampa e Pantanal.

Em relação a Mata Atlântica e áreas naturais não-florestais¹, seus limites atualmente ocupam 12,4% do país, distribuídas em 17 estados². Em consulta ao relatório “Atlas dos Remanescentes Florestais da Mata Atlântica”, no período de 1985 a 2019, a ação antrópica desmatou na ordem de 1,95 milhões de hectares (*ha*) desta cobertura florestal. O desflorestamento no período 2018 a 2019 foi de 14.502 *ha*, 27,2% superior ao período 2017 a 2018. Dos 17 estados monitorados entre 2018 a 2019, nove, registraram desmate nulo (< 100 *ha*), a saber: AL, CE, ES, GO, PB, PE, RJ, RN, SP (FUNDAÇÃO SOS MATA ATLÂNTICA & INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS [INPE], 2020a).

Destas regiões, destaca-se o estado do Ceará, com remanescentes da Mata Atlântica e áreas naturais não-florestais ocupando respectivamente 64.064 e 125.519 *ha*, ou 1,3% do estado. Da região, 61 municípios possuem resquícios da mata nativa, e portanto, em observação aos dados

¹ Várzeas, campos de altitude naturais, refúgios vegetacionais, dunas, restinga herbácea, apicum, banhado e campo úmido.

² Alagoas (AL), Bahia (BA), Ceará (CE), Espírito Santo (ES), Goiás (GO), Mato Grosso do Sul (MS), Minas Gerais (MG), Paraíba (PB), Paraná (PR), Pernambuco (PE), Piauí (PI), Rio de Janeiro (RJ), Rio Grande do Norte (RN), Rio Grande do Sul (RS), Santa Catarina (SC), São Paulo (SP) e Sergipe (SE).

de desmatamento se vê que, desde 2014 o estado registra desmatamento nulo ($< 100 \text{ ha}$) de Mata Atlântica, com aumento de 40% no desmate no período 2017 a 2018 e 2018 a 2019. Já em relação as áreas não-florestais, somente no período 2018 a 2019 foram desmatados 804 ha de restinga arbórea (FUNDAÇÃO SOS MATA ATLÂNTICA & INPE, 2020a).

Em face ao apresentado, o objetivo geral é analisar a relação entre a degradação ambiental e atividade econômica, preconizada pela hipótese da CKA, nos 61 municípios cearenses que apresentam remanescentes da Mata Atlântica e ecossistemas associados, considerando o indicador de desmatamento. Logo, busca-se saber se há uma relação no formato

de “U-invertido” entre o *PIBpc* e a área desmatada anual obtida pelo projeto “Aqui tem Mata?”, no período de 2011 a 2017, utilizando um modelo para dados em painel. Além da variável relacionada a atividade econômica, foram adicionados controles associados a: agropecuária, consumo, demografia, informações econômicas e fiscais, mercado de trabalho e aspectos geoambientais.

Embora diversas pesquisas estimem uma CKA para o Brasil a partir de diversos indicadores de desenvolvimento sustentável, nenhuma destas utilizou o desmatamento da Mata Atlântica e das áreas naturais não-florestais como *proxy* de degradação ambiental. Ademais, as análises empíricas da CKA tem como foco estados e municípios que possuem limites da Amazônia e Cerrado, ou seja, a literatura não apresenta evidências para o Ceará, no que tange os domínios da Caatinga e Mata Atlântica. Portanto, a presente pesquisa contribui com a literatura da CKA para o bioma Mata Atlântica.

Para alcançar esse objetivo, optou-se por dividir o artigo em cinco seções, incluindo essa introdução. A próxima dedica-se a expor a literatura empírica sobre a CKA com foco em desmatamento. A fonte e descrição dos dados, bem como o modelo empírico, são apresentados na terceira seção. Em seguida, são expostos, discutidos e analisados os resultados. Por fim, são apresentadas as conclusões.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 A literatura empírica sobre a CKA com foco em desmatamento

Para conduzir a revisão da literatura, optou-se por utilizar os descritores “*Environmental Kuznets Curve*” + “*Deforestation*” na base de dados eletrônica do *Google Scholar*. Os critérios de inclusão dos artigos foram: publicações em periódicos nacionais ou internacionais, nos idiomas inglês, português ou espanhol, com data de publicação a partir do ano de 2010, selecionando referências atualizadas e com predileção a estudos realizados em biomas brasileiros. A partir desses procedimentos de seleção, o conteúdo do artigo é avaliado considerando a área de estudo, período, metodologia e resultados.

Oliveira *et al.* (2011) investigaram o desmatamento sob a hipótese da CKA em municípios da Amazônia Legal no período de 2001 a 2006. Para tal, controlaram por indicadores relacionados a agropecuária, demografia, economia, extração florestal e mineral. O formato encontrado para CKA foi de “N-invertido”, assim, o desmatamento decresce nos níveis baixos de renda *pc*, passa por um período de aumento e volta a decrescer quando um elevado nível de renda *pc* é alcançado.

Ferreira e Coelho (2015) exploraram como os preços das *commodities* agrícolas, as políticas públicas de fiscalização e o crédito rural, têm afetado o desmatamento na região da Amazônia Legal durante os anos de 1999 a 2011. Os resultados obtidos levam a crer que os preços

das *commodities* e as políticas de crédito rural aumentam o desmatamento, ao passo que as políticas de fiscalização têm auxiliado na mitigação do desmatamento.

No contexto da discussão entre o desmatamento e desenvolvimento regional, Teixeira, Bertella e Almeida (2012) investigaram a relação entre os níveis de desmate e o crescimento da renda para 139 municípios do estado do MT. Utilizando dados para o ano de 2006 estimou-se uma CKA para o desmatamento *pc* em relação a renda *pc* e seu termo ao quadrado, além da: bovinocultura/área de pastagem, a densidade demográfica, a extração de madeira *pc* e efeitos espaciais. Os resultados encontrados trouxeram a inferência de que existe uma CKA na forma de “U-invertido”, porém ao se utilizar um termo cúbico para a renda, o crescimento econômico não estaria atrelado com o desmatamento dos municípios mato-grossenses.

Colusso *et al.* (2012) discutem a relação entre degradação ambiental e atividade econômica, sob a ótica da CKA, no Cerrado brasileiro, para o ano de 2008. A variável dependente é a área desmatada de Cerrado dos municípios do bioma e, como variáveis explicativas, o *PIBpc* e suas formas quadrática e cúbica, a densidade demográfica, a área plantada e o rebanho de bovinos. A hipótese testada resultou que, em um primeiro estágio, o crescimento da renda *pc* contribui para redução do desmatamento na região; porém, o aumento contínuo da renda faz com que o desmatamento aumente novamente.

Para o MATOPIBA³, Barros e Stege (2019) investigaram a existência de uma CKA da relação entre desmatamento e Índice de Desenvolvimento Humano (IDH) nos 337 municípios da região para o ano de 2010. A Análise Exploratória de Dados Espaciais (AEDE) aliada a econometria espacial corroborou o formato da CKA de “U-invertido”. O ponto de virada, em que o desenvolvimento atinge seu máximo é um IDH de 0,57, sendo que 28,18% dos municípios estão abaixo deste valor. Além disso, os controles relacionados ao avanço da fronteira agrícola induzem o desmatamento na região.

Apesar dos artigos supracitados abordarem de forma direta o desmatamento e sua relação com o crescimento econômico, existem outros (ALMEIDA & LOBATO, 2019; BIAGE & ALMEIDA, 2015; CARVALHO & ALMEIDA, 2010) que embora não tratem de forma explícita, explicam a emissão de gases estufa como uma consequência do desmatamento e que por sua vez são afetadas pelo crescimento econômico⁴.

Carvalho e Almeida (2010) testaram a hipótese da CKA em uma amostra de 187 países para o ano de 2004. A variável dependente são as emissões de CO_2 e como variáveis independentes – o *PIBpc* e suas formas quadráticas e cúbicas, as exportações *pc*, consumo de energia *pc* e uma *dummy* indicando os países signatários ao Protocolo de *Kyoto*. Os resultados corroboram o formato de “U-invertido”, ou seja, o crescimento econômico reduz o impacto ambiental das emissões, porém a adição do termo cúbico traz a curva no formato de “N”, demonstrando que elevados níveis de crescimento podem aumentar as emissões de CO_2 . Além disso, os países signatários ao Protocolo de *Kyoto* reduziram significativamente as emissões de poluentes.

Biage e Almeida (2015) avaliaram a hipótese da CKA aplicada a um formato de dados em painel, analisando as diferenças entre as emissões de CO_2 por países, em função do desenvolvimento socioeconômico. Como resultado, evidenciou-se relação entre *PIBpc* e a emissão

³ Atual fronteira agrícola brasileira, o MATOPIBA abrange o bioma Cerrado dos estados do Maranhão, Tocantins, Piauí e Bahia, e responde por grande parte da produção nacional de grãos e fibras (Empresa Brasileira de Agropecuária [EMBRAPA], 2020).

⁴ Almeida e Lobato (2019), Biage e Almeida (2015) corroboram que o incremento da atividade econômica eleva o desmatamento via setor agropecuário com: queimadas na agricultura, extração de recursos florestais, e áreas de pastagem, que reduzem a absorção do CO_2 , elevando as emissões de GEE.

de CO_2pc com CKA no formato de “N”. Além disso, os resultados demonstraram que o $PIBpc$ constitui a variável de menor impacto sobre as emissões de CO_2pc , e que o impacto ambiental cresce, essencialmente, pelo desenvolvimento das economias (desenvolvimento econômico, desenvolvimento social e qualidade de vida).

Almeida e Lobato (2019) propõem a discussão da CKA para a região Norte do Brasil no período de 2002 a 2015. Os autores relacionaram as emissões de CO_2 com o $PIBpc$, e apesar de não abordarem diretamente o desmatamento, levaram em consideração que a emissão de CO_2 é uma consequência de fatores como o próprio desmatamento, sendo tratada como variável *proxy*. A pesquisa concluiu que a CKA para a região Norte do Brasil apresenta formato de “U”, não corroborando com o formato tradicional da CKA.

Estudos relevantes para a construção do presente referencial, dizem respeito a avaliação do desmatamento, sem a necessidade de se testar a hipótese da CKA, a importância destes se dá na avaliação dos determinantes do desflorestamento em biomas brasileiros, ou da elaboração de indicadores de impacto ambiental, no sentido de identificar causas da degradação ambiental.

Sob outra perspectiva, Delazeri (2016) elencou as causas do desmatamento nos municípios do Arco Verde na Amazônia Legal entre 2008 e 2012. Com o resultado concluiu-se que a produção bovina tem maior incidência nos níveis de desmatamento dos 49 municípios que compõem a região e que a expansão da lavoura de soja não é significativa para explicar o desmatamento.

Em investigação sobre a ocorrência de desacoplamento de impacto ambiental do crescimento econômico brasileiro, Soares e Almeida (2018) agruparam 13 indicadores de pressão ambiental, monitorados a partir da década de 1990 a meados de 2014, nas dimensões⁵: água, atmosfera, biodiversidade, saneamento e terra. A pesquisa corroborou a hipótese de que o crescimento econômico ocasiona impacto ambiental, incluindo na biodiversidade que inclui a Mata Atlântica, porém com menor impacto na dimensão terra.

Castelo, Adami, Almeida e Almeida (2018) avaliaram políticas públicas ambientais do governo federal no combate ao desmatamento na Amazônia, a partir de um levantamento histórico da atividade por um período de 15 anos (2002 a 2016). A análise empírica realizada para o estado do Pará a partir de uma modelagem em painel mostrou que, dado o aumento de 100 cabeças bovinas, a área desmatada sofre um incremento de $0,2 km^2$, já a política ambiental foi bastante significativa para o período analisado.

Resumidamente, o Quadro 1 sintetiza os estudos que avaliam a relação entre crescimento econômico e desmatamento em biomas brasileiros sob a hipótese da CKA. Em geral, estes trabalhos empíricos consideram abordagens econométricas para dados em corte transversal (*cross-section*) ou no formato de dados em painel. Além disso, boa parte das pesquisas nacionais está concentrada nas regiões Centro-Oeste, Norte e nos municípios do MATOPIBA. Assim, por meio desta revisão, pode-se dizer que a presente pesquisa contribui com a literatura da CKA a partir da análise da relação entre a atividade econômica no estado do Ceará com o desmatamento dos remanescentes florestais da Mata Atlântica e áreas naturais não-florestais.

⁵ Água: demanda bioquímica de oxigênio (1990 a 2014), qualidade das águas das praias (1992 a 2012). Atmosfera: emissões de origem antrópica de GEE (1990 a 2014), consumo industrial de substâncias destruidoras da camada de ozônio (1992 a 2013). Biodiversidade (1992 a 2013): áreas terrestres protegidas, áreas de proteção marinha. Saneamento (1992 a 2011): acesso da população a água potável, esgotamento sanitário e ao serviço de coleta de lixo doméstico. Terra: desmatamento da Amazônia Legal (1990 a 2014), uso de fertilizante (1992 a 2013), uso da terra (1990 a 2011).

Quadro 1: Síntese dos trabalhos que considera a CKA e/ou o Desmatamento.

Autor(es)	Período	Método	Região	CKA	Desmate
Cross-Section					
Barros e Stege (2019)	2010		337 municípios do MATOPIBA	✓	✓
Carvalho e Almeida (2010)	2010	Econometria espacial	187 países	✓	-
Colusso <i>et al.</i> (2012)	2008		1.306 municípios que apresentam área de Cerrado no Brasil	✓	✓
Teixeira <i>et al.</i> (2012)	2006		Municípios do Mato Grosso	✓	✓
Dados em Pannel					
Almeida e Lobato (2018)	2002 a 2015		Estados da região norte brasileira	✓	-
Biage e Almeida (2015)	1991 a 2006		37 países	✓	-
Castelo <i>et al.</i> (2018)	2002 a 2016	Efeitos Fixos	Municípios do Pará	-	✓
Delazeri (2016)	2008 a 2012		Arco Verde (Amazônia Legal)	-	✓
Ferreira e Coelho (2015)	1999 a 2011		Estados da Amazônia Legal	✓	✓
Oliveira <i>et al.</i> (2011)	2001 a 2006	Econometria Espacial	Municípios da Amazônia Legal	✓	✓
Soares e Almeida (2018)	----	Construção de Indicadores	Brasil	-	✓

Fonte: Elaborada com base nos dados da pesquisa.

3. METODOLOGIA

3.1 Fonte de dados

Essa pesquisa analisa o desmatamento anual dos remanescentes florestais da Mata Atlântica e áreas naturais não-florestais – *DESM*, em *ha*. Esses dados são do projeto “Aqui tem Mata?”, aplicativo elaborado com dados do “Atlas da Mata Atlântica”, da Fundação SOS Mata Atlântica e INPE (2020b) que apresenta gráficos e mapas interativos com informação atualizada do estado de conservação das florestas, mangues e restingas em todos os 3.429 municípios brasileiros que apresentam resquícios de Mata Atlântica.

No levantamento das alterações na cobertura florestal, destacam-se duas limitações: a presença da cobertura de nuvens, que prejudica o processamento das imagens, podendo haver áreas não observadas; e, a limitação do mapeamento, que necessita de uma área mínima de 3 *ha*, tanto para a detecção das alterações florestais como na identificação dos remanescentes florestais. Para fins de monitoramento, áreas com desmatamento inferior a 3 *ha* são marcadas como indício de desmatamento e serão observadas novamente em novas versões dos relatórios (FUNDAÇÃO SOS MATA ATLÂNTICA & INPE, 2020a).

A consulta aplicada de 2011 a 20176 no estado do Ceará resultou em 61 municípios monitorados e distribuídos entre cinco das sete mesorregiões cearenses (Figura 1). A mesorregião 1 compreende 26 municípios monitorados (50,51% do total), seguida da mesorregião 2 com 20 municípios (28%) e mesorregiões 3, 5 e 7 conjuntamente com 14 municípios monitorados (21,49%).

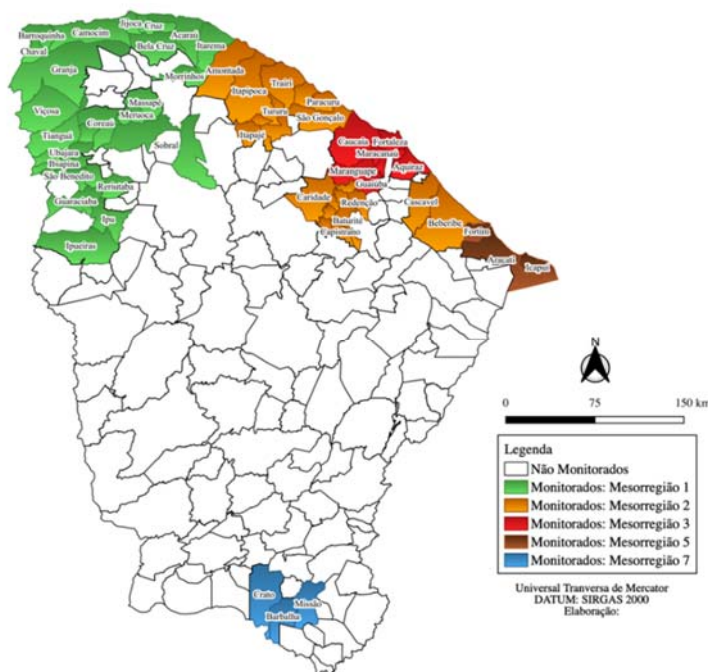
O processo de desmatamento pode se dar em função de diversas atividades, em especial das antrópicas. Com base nisso, utilizam-se de informações da agropecuária, consumo, demografia, econômicas e fiscais⁷, mercado de trabalho e aspectos geoambientais, provenientes de diversas fontes, como observado no Quadro 2.

As informações de economia, população e território são extraídas do IBGE, e incluem o *PIB_{pc}*, suas formas quadrática e cúbica, o Valor Adicionado Bruto da Agropecuária (*VAB_{pc}^{agro}*), o Valor da Produção Extrativista (*VPE*), a Densidade Populacional (*DPOP*) e o efetivo do rebanho bovino (*BOV*).

O número de trabalhadores efetivamente empregados nos setores da Agricultura, Pecuária e Serviços Relacionados e Produção Florestal (*Emp_{primário}*) são provenientes da Relação Anual de Informações Sociais (RAIS), do Ministério do Trabalho (MTB, 2017). Enquanto que o Consumo de Energia Elétrica Municipal (*CEpc*), é fornecido pela Enel (2017).

As informações fiscais como o Fundo de Participação Municipal (*FPM*), o Imposto sobre Circulação de Mercadorias e Serviços (*ICMS*) e Receita Tributária (*RT*) foram extraídas do Sistema de Informações Contábeis e Fiscais do Setor Público (SICONFI, 2017), banco de dados criado pela Secretaria do Tesouro Nacional (STN).

Figura 1: **Municípios cearenses monitorados pelo programa “Aqui tem Mata?”.**



Fonte: Elaborada com base nos dados da pesquisa.

6 Durante a elaboração desta pesquisa, não havia sido divulgado o *PIB* municipal referente ao ano de 2018, logo há uma limitação para o uso dos dados sobre o desmatamento para o ano de 2017.

7 Ambos indicadores deflacionados pelo Índice Geral de Preços – Disponibilidade Interna (IGP-DI), tomando-se por base o ano de 2015.

A Precipitação total anual observada (*PRECI*) é da Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME, 2020) e a delimitação da Região Semiárida (RSA) no Brasil são do Ministério da Integração (MI, 2018). Quanto as áreas protegidas, foram geradas variáveis binárias relacionadas a presença das Unidades de Conservação (*UC*) terrestres municipal, estadual e federal localizadas em território cearense (Ministério do Meio Ambiente [MMA], 2020).

Além disso, a base municipal conta com variáveis derivadas, construídas a partir do cruzamento de dois indicadores de diferentes fontes. Estas variáveis indicam valores integrais como proporção da população ou da área do município. Citam-se, informações econômicas, fiscais e de consumo de energia elétrica em valores *per capita*, já a bovinocultura e população são expressas em termos de área municipal.

Quadro 2: Descrição das variáveis e as fontes dos dados.

Variável	Descrição	Unidade	Fonte
<i>DESM</i>	Área desmatada dos remanescentes florestais da Mata Atlântica e áreas naturais não-florestais	ha	Fundação Mata Atlântica e INPE (2020b)
<i>PIBpc</i>	<i>PIB per capita</i> municipal	R\$ de 2015	IBGE (2017a)
$(PIBpc)^2$	<i>PIB per capita</i> municipal ao quadrado	-	Derivado do <i>PIBpc</i>
$(PIBpc)^3$	<i>PIB per capita</i> municipal ao cubo	-	
VAB_{pc}^{agro}	<i>VAB da agropecuária per capita</i>		IBGE (2017b)
<i>VPEpc</i>	Valor da produção extrativista de carvão vegetal, lenha e madeira em tora <i>per capita</i>	R\$ de 2015	IBGE (2017c)
<i>DPOP</i>	Razão da população absoluta e área ocupada por esse mesmo contingente populacional	hab/km ²	IBGE (2017d)
<i>BOV</i>	Razão do efetivo do rebanho de bovinos pela área municipal	contagem/km ²	IBGE (2020)
<i>Emp.primário</i>	Empregos formais nos setores da Agricultura, Pecuária e Serviços Relacionados + Produção Florestal	contagem	MTB (2017)
<i>CEpc</i>	Consumo municipal de energia elétrica <i>per capita</i>	MWh pc	Enel (2017)
<i>FPMpc</i>	Fundo de Participação Municipal <i>per capita</i>		
<i>ICMSpc</i>	ICMS <i>per capita</i>	R\$ de 2015	SICONFI (2017)
<i>RTpc</i>	Receita Tributária <i>per capita</i>		
<i>PRECI</i>	Precipitação total anual observada	mm	FUNCEME (2020)
<i>RSA</i>	1 se município pertence ao Semiárido, 0 caso contrário (c.c)		MI (2018)
<i>UCmunicipal</i>	1 se o município possui <i>UC</i> municipal, 0 c.c	variável binária	
<i>UCestadual</i>	1 se o município possui <i>UC</i> estadual, 0 c.c		MMA (2020)
<i>UCfederal</i>	1 se o município possui <i>UC</i> federal, 0 c.c		

Fonte: Elaborada com base nos dados da pesquisa.

3.2 Modelo Empírico

Na estimativa da CKA aplicada ao desmatamento dos remanescentes florestais e áreas naturais não-florestais no Ceará, considerou-se a especificação do modelo que inclui a variável *PIBpc* na forma quadrática e cúbica, na seguinte forma:

$$DESM_{it} = \beta_0 + \beta_1 PIBpc_{i,t-1} + \beta_2 (PIBpc)_{i,t-1}^2 + \beta_3 (PIBpc)_{i,t-1}^3 + \beta_k X_{i,t-1} + \varepsilon_{it} \quad (1)$$

Em que, $DESM_{it}$ é o desmatamento anual para cada um dos i municípios monitorados pelo programa “Aqui tem Mata?”, sendo $i = 1, \dots, 61$. O subscrito t para o conjunto de dados corresponde ao ano observado $t = 2011, \dots, 2017$, já o $PIBpc_{i,t-1}$ denota o nível da atividade econômica do i -ésimo município no ano anterior e $X_{i,t-1}$ corresponde ao conjunto de variáveis explicativas adicionais defasadas em um ano (Quadro 2). Os dados compõem um painel balanceado com 427 observações.

A partir disso, o formato da CKA está relacionado ao sinal e significância apresentados pelos coeficientes $(\beta_1, \beta_2, \beta_3)$. Uma condição suficiente para que a CKA apresente formato linear se dá quando $\beta_1 > 0$ ou $\beta_1 < 0$, enquanto que $\beta_2 = \beta_3 = 0$. No caso de $\beta_1 > 0$, um aumento no $PIBpc$ é linearmente relacionado ao desmatamento. Para o formato de “U-invertido” é suficiente que $\beta_1 > 0, \beta_2 < 0$ e $\beta_3 = 0$, já para o formato de “U”, $\beta_1 < 0, \beta_2 > 0$ e $\beta_3 = 0$. Por fim, nos casos em que $\beta_1 > 0, \beta_2 < 0$ e $\beta_3 > 0$ ou $\beta_1 < 0, \beta_2 > 0$ e $\beta_3 < 0$, a curva assume forma de “N” e “N-invertido” respectivamente.

Note que, a variável dependente, o desmatamento anual dos remanescentes florestais de Mata Atlântica e áreas naturais não-florestais pode ser considerada como uma variável censurada, pois é razoável supor que o desmatamento não assume valores negativos, assim não observamos a variável de interesse se ela está abaixo do zero. Além disso, a análise destes dados revela que, das 427 observações para o desmatamento, 347 são observações nulas (desmate < 3 ha), ou seja, há um excesso de zeros e não a ausência de informações.

Ao desconsiderar tais fatos (a censura e o excesso de zeros) as estimativas tradicionais de Mínimos Quadrados Ordinários serão inconsistentes, pois os pressupostos do modelo linear clássico são violados. Um modelo que se aplica bem a estes problemas é modelo *tobit*, proposto inicialmente por Tobin (1958), adequado quando a variável dependente é censurada ou truncada, além destes casos se adequa quando a variável dependente assume um número excessivo de zeros⁸ (CALZOLARI, MAGAZZINI & MEALLI, 2001). Assim, o modelo *tobit* pode ser expresso na seguinte forma:

$$Y_{it} = \begin{cases} 0, & \text{se } Y_{it}^* \leq 0 \\ Y_{it}^*, & \text{se } Y_{it}^* > 0 \end{cases} \quad (2)$$

Em que, $Y_{it}^* = X'_{it}\beta + \varepsilon_{it}$ é observado apenas se estritamente positivo, X_{it} representa o vetor de variáveis explicativas e o termo $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ *i.i.d* e independente de X_{it} , com $i = 1, \dots, N$ e $t = 1, \dots, T$. A função densidade de probabilidade quando Y é observado é (AMEMIYA, 1985):

$$f(Y_{it}|X_{it}; \theta) = \begin{cases} 0, & \text{se } Y_{it} < 0 \\ \Phi(-X'_{it}\beta/\sigma_\varepsilon), & \text{se } Y_{it} = 0 \\ \phi((Y_{it} - X'_{it}\beta)/\sigma_\varepsilon), & \text{se } Y_{it} > 0 \end{cases} \quad (3)$$

Onde, Φ é a função de distribuição acumulada e ϕ é a função densidade de probabilidade da distribuição normal padrão. Ao considerar o modelo para dados em painel, o termo de erro ε_{it} pode ser decomposto em (CALZOLARI, MAGAZZINI & MEALLI, 2001):

$$\varepsilon_{it} = \alpha_i + \lambda_t + u_{it} \quad (4)$$

Onde α_i são os efeitos individuais (características não observáveis específicas da unidade i que são consideradas constantes ao longo do tempo) interpretados como parâmetros fixos ou

⁸ Não confundir com modelos para dados de contagem inflacionados de zeros: *Poisson* Inflacionado de Zeros (*Zero-Inflated Poisson*, *ZIP*) ou Binomial Negativo Inflacionado de Zeros (*Zero-Inflated Negative Binomial*, *ZINB*).

como variáveis aleatórias, já λ_t é o efeito do tempo (características não observáveis no período t , constante para todas as unidades de corte transversal na amostra) e u_{it} é um termo aleatório que varia no tempo e com os indivíduos, além disso u_{it} não está correlacionada ao longo do tempo (CALZOLARI, MAGAZZINI & MEALLI, 2001).

Modelos de regressão *tobit* com dados em painel podem assumir duas formas, sendo efeitos fixos ou efeitos aleatórios. Segundo Cameron e Trivedi (2005), em painéis menores ($T < 8$) o estimador de efeitos fixos não se mostram consistentes, assim o modelo de efeitos aleatórios é o mais indicado:

$$Y_{it}^* = \alpha_i + X'_{it}\beta + u_{it} \quad (5)$$

Em que, $\alpha_i \sim N(0, \sigma_\alpha^2)$ *i.i.d.*, $u_{it} \sim N(0, \sigma_u^2)$ *i.i.d.*, com u_{it} independente de α_i , assume-se que da expressão (4) o termo $\lambda_t = 0 \forall t$. A equação (5) pode ser reescrita como (CALZOLARI, MAGAZZINI & MEALLI, 2001):

$$Y_{it}^* = \sigma_\alpha \alpha_i + X'_{it}\beta + \sigma_u u_{it} \quad (6)$$

Onde $\alpha_i \sim N(0,1)$ *i. i. d* independente de u_{it} , com $u_{it} \sim N(0,1)$ *i. i. d*. Devido ao efeito individual, as observações sobre a variável dependente para cada indivíduo i são correlacionadas. No entanto, atrelado ao efeito individual α_i , a função de densidade conjunta condicional pode ser escrita como (GOURIEROUX & MONFORT, 1993):

$$f(Y_i|X_i, \alpha_i; \theta) = \prod_{t:Y_{it}>0} \frac{1}{\sigma_u} \phi\left(\frac{Y_{it} - X'_{it}\beta - \sigma_\alpha \alpha_i}{\sigma_u}\right) * \prod_{t:Y_{it}=0} \Phi\left(\frac{-X'_{it}\beta - \sigma_\alpha \alpha_i}{\sigma_u}\right) \quad (7)$$

Φ e ϕ são a função de distribuição acumulada e a função densidade de probabilidade com distribuição $N(0,1)$. Como os efeitos individuais α_i não são observáveis, a equação (6) não pode ser usada nas inferências. Para obter a máxima verossimilhança incondicional é preciso integrar o efeito individual α_i :

$$f(Y_i|X_i; \theta) = \int f(Y_i|X_i, \alpha_i; \theta) dP^\alpha(\alpha) \quad (8)$$

Uma solução satisfatória para a integral (7) é o procedimento via integração numérica. Alternativamente, pode-se aproximar a integral (7) por meio de simulações replicadas, obtendo-se assim uma função de verossimilhança simulada a ser maximizada afim de obter um estimador simulado de máxima verossimilhança (CALZOLARI, MAGAZZINI & MEALLI, 2001). Por fim, como análises adicionais a regressão *tobit* com efeitos aleatórios serão testadas CKA com especificações da variável *PIBpc* na forma quadrática e cúbica para modelos *tobit* com dados empilhados ou *pooled tobit* (modelos-base).

4 RESULTADOS E DISCUSSÕES

Os resultados são apresentados e discutidos em três subseções que concentram as análises dos resultados descritivos, seguido pelas áreas remanescentes e desmatadas e, pro fim, são apresentados os resultados econométricos a partir das especificações da CKA com termo do *PIBpc* nas formas quadrática e cúbica, estimadas por regressões *pooled tobit* e *tobit* com efeitos aleatórios.

4.1 Análise Descritivas dos Dados

A Tabela 1 apresenta algumas estatísticas descritivas para o conjunto de dados. Em linhas gerais, todas as variáveis apresentaram variação percentual positiva no intervalo de 2011 a 2017,

a exceção da bovinocultura, do percentual de empregos no setor primário da economia em relação ao total de empregos formais e a precipitação.

A análise do desmatamento, revela média anual de 6,93 *ha*/município, com máxima extensão desmatada de 375 *ha* no município de Trairi (litoral oeste do estado do Ceará). Destas observações, grande parte (81%) se referem a desmatamento abaixo dos 3 *ha*, logo são 347 observações nulas. Apesar de tal fato, o desmatamento nestas regiões pode estar ocorrendo pelo “efeito formiga”, desmatamentos pequenos o suficiente para que o satélite não capte sua presença.

Das informações econômicas, destaque para o *PIB_{pc}*, com média de R\$ 10 mil, mínimo de R\$ 2,7 mil para a cidade de Tururu e máximo de R\$ 71 mil em São Gonçalo do Amarante (SGA). Para o *VAB_{pc}^{agro}*, a média foi de R\$ 1,1 mil, com mínimo registro na cidade de Fortaleza e máximo valor em Missão Velha. Já o *VPE_{pc}* tem média de R\$ 6,87, com máximo valor *pc* em Frecheirinha (R\$ 122,48).

No tocante a *DPOP*, a média é de 263 habitantes a cada *km²*, com adensamento máximo na cidade de Fortaleza (capital do estado do Ceará) e mínimo no município de Granja. A bovinocultura tem média de 14,60 cabeças/*km²*, com registros mínimo e máximo nas cidades serranas de Guaramiranga, com 2,50 cabeças/*km²* e Maranguape, com 34,41 cabeças/*km²*, localizada na Região Metropolitana de Fortaleza – RMF.

Quanto aos empregos no setor primário da economia, verifica-se uma baixa participação em relação ao total de empregos formais, com maior proporção (> 45%) no município de Icapuí (litoral leste). Em relação ao *CE_{pc}*, a média é de 0,98 *MWh*, com consumo mínimo na cidade de Granja e máximo em SGA. Por fim, para as informações fiscais, as médias do *FPM_{pc}*, *ICMS_{pc}* e *RT_{pc}* foram respectivamente, R\$ 560,65, R\$ 212,88 e R\$ 111,24. Destaque para o município de SGA, que detém as maiores métricas relacionadas ao *PIB_{pc}*, *CE_{pc}*, *ICMS_{pc}* e *RT_{pc}*, possivelmente pela presença do Complexo Industrial e Portuário do Pecém (CIPP) na região.

A matriz de correlação é apresentada no Apêndice A, a partir da qual se observa uma alta correlação entre *PIB_{pc}* e *ICMS_{pc}* e entre *PIB_{pc}* e *RT_{pc}*, respectivamente, 0,806 e 0,842. Em função disso, optou-se por considerar apenas o *PIB_{pc}* no modelo econométrico.

Tabela 1: Estatísticas descritivas, 2011-2017.

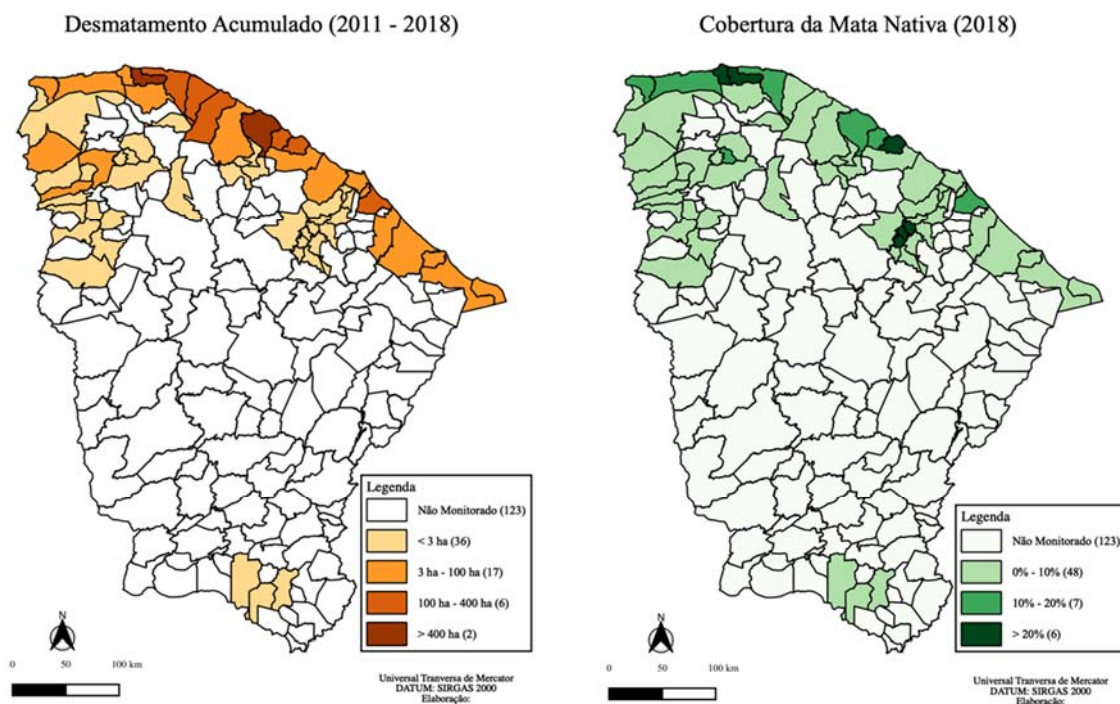
	Média	Desvio-Padrão	Mínimo	Máximo	$\Delta\%$ (2011-2017)
<i>DESM</i>	6,93	31,08	0,00	375,00	58,68%
<i>PIB_{pc}</i>	10056,60	8453,16	2751,99	71026,16	153,27%
<i>VAB_{pc}^{agro}</i>	1148,80	1102,73	11,23	8899,82	58,00%
<i>VPE_{pc}</i>	6,87	11,65	0,00	122,48	16,76%
<i>DPOP</i>	263,00	1052,56	19,61	8343,06	6,82%
<i>BOV</i>	14,60	6,76	2,50	34,41	-22,50%
<i>Emp_{primário}</i>	5,07	9,72	0,00	71,41	-1,60%
<i>CE_{pc}</i>	0,98	0,91	0,32	6,87	16,47%
<i>FPM_{pc}</i>	560,65	296,60	0,00	2681,13	153,20%
<i>ICMS_{pc}</i>	212,88	185,09	0,00	1283,49	228,47%
<i>RT_{pc}</i>	111,24	196,27	0,00	2245,39	308,86%
<i>PRECI</i>	916,18	381,01	19,00	2382,00	-30,95%

Fonte: Elaborado pelos autores.

4.2 Análise sobre o desmatamento acumulado e mata nativa remanescente

A Figura 2 apresenta a distribuição espacial do desmatamento acumulado de Mata Atlântica e ecossistemas associados no período de 2011 a 2018, bem como a proporção destes domínios em relação a área municipal para o ano de 2018 (ano mais recente para as duas informações). Para o desmatamento, cores mais fortes evidenciam maior degradação, neste sentido destacam-se os municípios de Cruz e Trairi que respondem a 45% do total desmatado no intervalo de oito anos.

Figura 2: Desmatamento acumulado e cobertura da mata nativa no estado do Ceará.



Fonte: Elaborado pelos autores.

Além disso, o desmatamento está localizado sobretudo no litoral, em especial nos municípios da porção oeste do estado (Acará, Amontada, Cruz, Itarema, Paracuru, Paraipaba e Trairi). O recente desmatamento no litoral oeste pode se dar pela configuração destes locais como novos destinos turísticos, com o loteamento das terras para a construção de casas de veraneios e *resorts*, em face ao baixo custo da terra em alguns locais. Por outro lado, identificam-se 36 municípios monitorados com desmatamento inferior a 3 ha. Apesar de tal fato, é preciso levar em consideração o “efeito formiga” imposto pelo avanço das moradias e expansão da infraestrutura urbana.

Para a cobertura da mata nativa, cores mais fortes evidenciam maior proporção das áreas de Mata Atlântica e áreas naturais não florestais em relação a extensão territorial. Assim, seis municípios possuem mais que 20% do território coberto por mata nativa, sendo eles: Cruz, Guaramiranga, Jijoca de Jericoacoara, Mulungu, Pacoti e Paracuru. Porém, nota-se que grande parte dos municípios monitorados encontra-se na faixa de mata nativa abaixo de 10% do território municipal.

4.3 Estimativas para a Curva de Kuznets Ambiental

Em relação a CKA, as versões na forma quadrática e cúbica foram estimadas por modelos de regressão *pooled tobit* e *tobit* com efeitos aleatórios. A Tabela 2 apresenta estes resultados, dentre os modelo-base (*pooled tobit*) e conforme o Critério de Informação de Akaike (AIC), o modelo com melhor ajuste (1093,693 contra 1090,122) é aquele que incorpora a relação cúbica entre o desmatamento e *PIBpc*.

Para tal modelo, são significativas as variáveis *DPOP*, *BOV*, *CEpc*, *FPMpc* e as *dummies* para o semiárido e *UC* federais. A densidade populacional mostra relação negativa com a área desmatada anual, denotando que municípios com grandes adensamentos populacionais desmatam menos áreas de mata nativa. Mesmo com o resultado inesperado, é possível notar que na literatura científica, não há um consenso sobre o sinal obtido para esta variável, pois algumas pesquisas encontraram relação positiva (CROPPER & GRIFFITHS, 1994), negativa (TEIXEIRA *et al.*, 2012) ou até mesmo não observaram relação (OLIVEIRA *et al.*, 2011; COLUSSO *et al.*, 2012) entre desmatamento e densidade populacional.

A pecuária bovina apresentou sinal inverso ao esperado (positivo), indicando que a presença desta atividade agropecuária contribui para um menor desmatamento. Colusso *et al.* (2012) corroboram a relação negativa entre o rebanho bovino e desmatamento nos 1.306 municípios que apresentam área de Cerrado no Brasil para o ano de 2008. Por outro lado, estes resultados contradizem os encontrados por Oliveira *et al.* (2011) e em Barros e Stege (2019).

O incremento do consumo de energia elétrica *pc* é negativamente relacionado ao desmatamento anual que se estende entre os 61 municípios cearenses analisados. Embora as pesquisas que avaliem a hipótese da CKA não abordem diretamente a relação entre consumo de energia e desmatamento, o fazem para as emissões de CO_2 , uma *proxy* para o desmatamento. Como consequência, os resultados obtidos indicam que as emissões de CO_2pc aumentam quando o *CEpc* se eleva (BIAGE & ALMEIDA, 2015; CARVALHO & ALMEIDA, 2010) conclusão contrária ao resultado desta pesquisa.

O *FPMpc* anterior está negativamente relacionado ao desmatamento, isto é, municípios que recebem maior repasse deste recurso desmatam menos em relação aos demais, o que está de acordo com o esperado. A transferência deste recurso aos municípios é atrelada ao fator população e ao *PIBpc*, assim cidades mais populosas e com alta renda *pc* recebem maior repasse. Por consequência, são mais desenvolvidas e possuem despesas significativas em diversos serviços ofertados à população, como educação e saúde, serviços sem relação direta ao setor florestal. Além disso, é razoável supor que áreas mais desenvolvidas dispõem de menos zonas verdes para desmatar.

As *dummies* geoambientais significativas foram a região semiárida e as *UC* federais. Os coeficientes apresentaram sinal negativo, levando a crer que os municípios inseridos em regiões de clima semiárido desmatam menos dos remanescentes florestais e não-florestais da Mata Atlântica em relação a aquelas com clima tropical úmido. Já, os municípios que possuem *UC* federais desmatam menos, isto é, a presença destas áreas de proteção ambiental não eliminam por si só a atividade do desmatamento nestas regiões, sobretudo o desmate ilegal.

Em relação a CKA, o fato do (*PIBpc*)³ ser positivo sugere que o desmatamento da mata nativa remanescente aumenta quando o *PIBpc* aumenta, no Ceará. Isto decorre do formato de “N” da CKA, em que os incrementos adicionais na atividade econômica (*PIBpc*) intensifica o desmatamento de Mata Atlântica. Os pontos de inflexão estimados são de R\$ 22 mil e R\$ 60 mil, sendo que o máximo local (R\$ 22 mil) contempla 91,8% dos municípios (responsável por 87% da

área desmatada), que pertencem a parte ascendente da curva. Quatro municípios estão na região de R\$ 22 mil $\leq PIBpc \leq$ R\$ 60, representando pouco mais de 11% do desmatamento. E, por fim, na região de inclinação ascendente, com $PIBpc \geq$ R\$ 60 mil encontra-se o município de SGA.

Analisando os resultados dos modelos *pooled tobit versus tobit* com efeitos aleatórios, verifica-se que o teste da razão de verossimilhança (*Likelihood Ratio, LR*) para modelos *tobit* em dados em painel rejeitam $H_0: \sigma_\alpha = 0$, ou seja, deve-se analisar os resultados da regressão *tobit* com efeitos aleatórios. Observe pelo Critério de Informação de Akaike (*AIC*) que os modelos *tobit* com efeitos aleatórios apresentam melhor performance em relação os modelos-base (*pooled tobit*). Quanto ao modelo, o que apresenta a relação cúbica entre o desmatamento e *PIBpc* tem o melhor ajuste (1022,569 contra 1018,819).

Em termos de efeito, verifica-se que a pressão demográfica exercida pela densidade populacional se relaciona negativamente com o desmatamento, ou seja, incrementos na *DPOP* suavizam a atividade do desmatamento. No entanto, esperava-se um efeito positivo dessa relação, entretanto, é possível que devido a expansão inicial da malha urbana, algumas áreas verdes deem lugar as produções imobiliárias e obras de infraestrutura urbana. Assim, com a baixa disponibilidade das zonas verdes, deve-se preservar os pequenos fragmentos florestais urbanos.

Quanto ao rebanho bovino, o efeito é negativamente relacionado a área desmatada anual nos municípios monitorados. A intuição para a escolha desta variável no modelo é a de que, com a expansão das áreas de pasto, ocorreriam alterações na cobertura florestal. Porém, observando a produção nacional, se vê que o estado do Ceará responde a pouco mais de 1% do mercado brasileiro, em contrapartida os estados de Goiás e de Mato Grosso responde a 10,6% e 14% respectivamente da produção bovina em referência ao ano de 2018 (IBGE, 2020), levantando suspeitas de que a atividade pecuária não é bem difundida no Ceará.

Além deste fato, os dados para a bovinocultura cearense revelam uma redução gradual da atividade no estado. Acrescenta-se ainda, que a pecuária bovina pode estar sendo praticada principalmente no domínio morfoclimático da Caatinga cearense, logo a bovinocultura não estaria diretamente relacionada ao desmatamento anual nos domínios da Mata Atlântica e ecossistemas associados.

Tabela 2: Resultados sobre os desmatamento dos remanescentes florestais da Mata Atlântica e áreas naturais não-florestais no Ceará, 2011-2017.

Modelo	Pooled Tobit		Efeitos Aleatórios Tobit	
	(<i>PIBpc</i>) ²	(<i>PIBpc</i>) ³	(<i>PIBpc</i>) ²	(<i>PIBpc</i>) ³
Constante	119,027*	77,937	52,377	6,986
<i>PIBpc</i>	0,017*	0,025*	0,012*	0,021*
(<i>PIBpc</i>) ²	-3,00e ⁻⁰⁷ *	-7,78e ⁻⁰⁷ *	-2,03e ⁻⁰⁷ *	-6,27e ⁻⁰⁷ *
(<i>PIBpc</i>) ³		6,32e ⁻¹² *		5,10e ⁻¹² *
<i>VAB^{agro}</i>	-0,014	-0,015	-0,018	-0,018
<i>VPEpc</i>	-0,394	-0,153	0,312	0,594
<i>DPOP</i>	-0,032*	-0,031*	-0,028*	-0,028*
<i>BOV</i>	-9,303*	-9,093*	-7,382*	-7,126*
<i>Emp.primário</i>	0,806	0,713	0,881	0,797
<i>CEpc</i>	-30,776*	-27,688*	-16,143	-15,455
<i>FPMpc</i>	-0,111*	-0,115*	-0,046	-0,063
<i>PRECI</i>	0,019	0,021	0,018	0,021
<i>RSA</i>	-95,389*	-95,859*	-82,405*	-77,345
<i>UCmunicipal</i>	-37,481	-40,106	8,605	4,369
<i>UCestadual</i>	14,326	12,964	20,110	19,179
<i>UCfederal</i>	-37,698*	-36,512*	-33,395	-33,896
	(12,977)	(12,898)	(22,782)	(22,575)

<i>/siama</i>	76,303	75,718		
<i>/siama α</i>			62,537*	61,792*
<i>/siama u</i>			48,907*	48,326*
LR			73,120*	73,300*
AIC	1093,693	1090,122	1022,569	1018,819
Pseudo R ²	0,070	0,087		
Nº Obs.	366	366	366	366
Ponto de inflexão	28333,333	Máx: 21923,520	29556,650	Máx: 23458,169
Forma Funcional	“U-invertido”	“N”	“U-invertido”	“N”

Fonte: Elaborada com base nos dados da pesquisa.

Notas: Erros-padrão robusto a heterocedasticidade entre parênteses. * denota a significância ao nível de 5%.

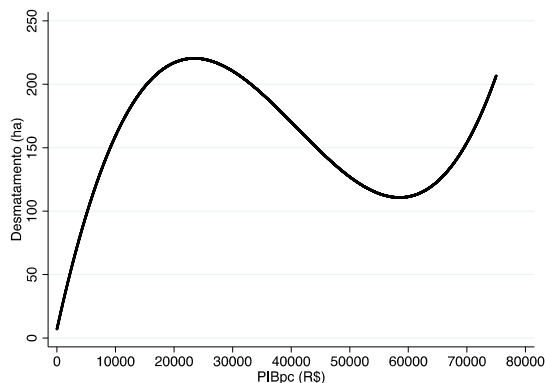
Como o $(PIBpc)^3$ é positivo, a CKA apresenta uma forma de “N”, o que implica que a redução do desmatamento, quando a atividade econômica está crescendo, é transitório (até a parte de “U-invertido), pois a medida que ocorrem aumentos adicionais no *PIBpc*, o desmatamento aumenta, como ilustrado na Figura 3. Os pontos de inflexão encontrados foram de R\$ 23 mil e R\$ 58 mil. O ponto de máximo local do modelo *pooled tobit* (R\$ 23 mil), responde por 91,8% dos municípios cearenses da amostra, os quais são responsáveis por 87% da área desmatada. Ao comparar o *PIBpc* médio dos municípios ao máximo local (R\$ 10 mil contra R\$ 23 mil) se observa que no curto prazo estas regiões ainda estarão na parte ascendente da curva e, por isso, ainda é possível que estes municípios adotem estratégias unilaterais (ou até de forma coordenada) para achatar a curva do desmatamento e, com isso, preservar as áreas remanescentes florestais da Mata Atlântica no Ceará, assim como as áreas naturais não-florestais.

Na parte descendente da curva, que vai $R\$ 23 \text{ mil} \leq PIBpc \leq R\$ 58 \text{ mil}$, figuram quatro municípios – Fortaleza, Aquiraz, Eusébio e Maracanaú (os três últimos na RMF), que respondem a pouco mais de 11% do desmatamento. Após o ponto de mínimo local ($PIBpc \geq R\$ 58 \text{ mil}$), na parte ascendente da curva, situa-se o município de São Gonçalo do Amarante, importante região econômica cearense por contemplar o CIPP e, ainda, por ter instalações portuárias, siderúrgicas, termelétrica, parques eólicos e a Zona de Processamento de Exportação (ZPE).

A partir da forma de “N” da CKA pode-se inferir que o contínuo aumento da atividade econômica municipal, aferido pelo *PIBpc*, afeta de forma positiva o desmatamento. E, como este voltou a crescer, não se pode afirmar que os níveis de desmatamento atingiram um máximo absoluto na região de estudo. O resultado é similar ao de Colusso *et al.* (2012) que observaram uma CKA em formato de “N” para a relação entre crescimento econômico e degradação ambiental no bioma Cerrado. Entretanto, as diferentes relações entre desmatamento e renda *pc*, estabelecidas pelo formato de “U-invertido” (FERREIRA & COELHO, 2015; BARROS & STEGE, 2019) e “N-invertido” (OLIVEIRA *et al.*, 2011; RODRIGUES, CUNHA, BRITO & PIRES, 2016) levam a conclusão de que a hipótese de “U-invertido” para o desmatamento não consegue se consolidar na literatura como um fato estilizado.

Por fim, quanto a hipótese da CKA, pode-se dizer que ela é verificada no formato de “U-invertido” apenas nos modelos com especificação quadrática do *PIBpc*. Ao incorporar o termo cúbico do *PIBpc* a CKA se configura na forma de “N”, rompendo o paradigma de que a atividade econômica gera por si só uma proteção automática ao meio ambiente, vis a vis ao desenvolvimento sustentável. Pelo contrário, as evidências encontradas por esta pesquisa apontam para a existência de um movimento cíclico na CKA nas áreas remanescentes florestais da Mata Atlântica e áreas naturais não-florestais no Ceará.

Figura 03. Curva de Kuznets Ambiental para o desmatamento dos remanescentes florestais da Mata Atlântica e áreas naturais não-florestais no Ceará, 2011-2017.



Fonte: Elaborada com base nos dados da pesquisa.

5 CONCLUSÕES

Nesta pesquisa investigou-se a hipótese de “U-invertido” para a CKA entre o *PIBpc* e desmatamento anual dos remanescentes florestais e não-florestais da Mata Atlântica em uma amostra de 61 municípios cearenses monitorados pelo programa “Aqui tem Mata?” no período de 2011 a 2017.

As causas do desmatamento são condicionadas a diversas atividades, em especial as antrópicas. Para contemplar estas relações, utilizam-se as características associadas a agropecuária, consumo, demografia, informações econômicas e fiscais, mercado de trabalho e aspectos geoambientais. Compreender estas relações pode auxiliar no entendimento do processo de desenvolvimento sustentável da região, desacelerando o desmatamento em domínios da Mata Atlântica no Ceará.

A relação entre desmatamento e atividade econômica é verificada na forma de “N”. Na fase ascendente da CKA ($PIBpc \leq R\$ 23 \text{ mil}$), o desmatamento é crescente para baixos níveis de *PIBpc*. Por outro lado, na região de R\$ 23 mil a R\$ 58 mil de *PIBpc*, o desmatamento diminui à medida que o *PIBpc* se eleva, tornando-se novamente crescente para valores de *PIBpc* superior a R\$ 58 mil. Outro aspecto relevante, é que 91,8% dos municípios encontra-se na primeira fase ascendente da curva, já o *PIBpc* médio do Estado é de R\$ 10 mil, o que sugere que no curto prazo, grande parte destes municípios não atingirão a faixa intermediária da CKA ($R\$ 23 \text{ mil} \leq PIBpc \leq R\$ 58 \text{ mil}$), onde o desmatamento diminui com o aumento da atividade econômica. As análises da CKA, indicam que o nível de *PIBpc* é um fator crucial para o desmatamento.

Em relação a pressão demográfica exercida pela densidade populacional e pela bovinocultura, verifica-se que elas reduzem o desflorestamento. Todavia, vale ressaltar que a pecuária bovina não é tão expressiva no Ceará, e que a expansão das áreas de pasto pode se dar nos domínios da Caatinga, não tendo relação direta com a Mata Atlântica e suas formas naturais não-florestais. Para a densidade populacional, o efeito pode ser associado a preservação dos fragmentos florestais em áreas urbanas e limítrofes, em virtude da baixa disponibilidade de zonas verdes na região.

Em face a estes resultados, o formato em “N” da CKA revela um comportamento municipal, em que o desenvolvimento econômico não estaria alinhado ao desenvolvimento sustentável dos recursos florestais, pois apenas quatro cidades estariam na parte intermediária da CKA, ao passo que, outras 56 cidades ainda estariam na primeira fase da curva, com intensa

relação entre atividade econômica e área desmatada. Portanto, conclui-se que o aumento do *PIBpc* por si só não é um fator que promova uma melhoria na qualidade ambiental.

A criação de novas tecnologias, assim como os esforços unilaterais dos municípios possibilitam a preservação os remanescentes florestais e áreas naturais não-florestais da Mata Atlântica, reduzindo assim o desmatamento na região, em especial o desmate ilegal. Vale ressaltar que o monitoramento destas áreas pode auxiliar no desenho de políticas públicas destinadas ao uso sustentável dos recursos naturais e redução dos níveis de desmatamento, pois as florestas e ecossistemas associados ofertam diversos serviços ambientais e, além disso, a proteção destes *habitats* selvagens podem prevenir a transmissão de zoonoses, por conta da interação mais direta entre seres humanos e animais promovida pela supressão ambiental.

6 REFERÊNCIAS

ALMEIDA, M. G.; LOBATO, T. C. A curva de Kuznets ambiental para a região norte do Brasil entre os anos de 2002 a 2015. **Economia & Região**, v. 7, n. 1, p. 7-25, 2019.

AMEMIYA, T. **Advanced econometrics**. Harvard University Press, Cambridge, MA, 1985.

BARROS, P. H. B.; STEGE, A. L. Deforestation and human development in the Brazilian agricultural frontier: an environmental Kuznets curve for MATOPIBA. **Revista Brasileira de Estudos Regionais e Urbanos**, v. 13, n. 2, p. 161-182, 2019.

BIAGE, M.; ALMEIDA, H. J. F. Desenvolvimento e impacto ambiental: uma análise da curva ambiental de Kuznets. **Pesquisa e planejamento econômico**, v. 45, n. 3, p. 505-556, 2015.

CALZOLARI, G.; MAGAZZINI, L.; MEALLI, F. Simulation-based estimation of Tobit model with random effects. **Econometric Studies: a Festschrift in Honour of Joachim Frohn**, p. 349-369, 2001.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: methods and applications**. New York: Cambridge University Press, 2005.

CARVALHO, T. S.; ALMEIDA, E. S. A hipótese da curva de Kuznets ambiental global: uma perspectiva econométrico-espacial. **Estudos Econômicos**, v. 40, n. 3, p. 587-615, 2010.

CASTELO, T. B.; ADAMI, M.; ALMEIDA, C.; ALMEIDA, O. T. Governos e mudanças nas políticas de combate ao desmatamento na Amazônia. **Revista Iberoamericana de Economía Ecológica**, v. 28, n. 1, p. 125-148, 2018.

COLUSSO, M. V. S.; PARRÉ, J. L.; ALMEIDA, E. S. Degradação ambiental e crescimento econômico: a curva de Kuznets ambiental para o cerrado. **Revista de Economia e Agronegócio**, v. 10, p. 335-357, 2012.

CROPPER, M.; GRIFFITHS, C. The Interaction of Population Growth and Environmental Quality. **The American Economic Review**, v. 84, n. 2, p. 250-254, 1995.

DELAZERI, L. M. M. Determinantes do desmatamento nos municípios do arco verde - Amazônia Legal: uma abordagem econométrica. **Economia Ensaios**, v. 30, n. 2, p. 11-34, 2016.

EMBRAPA. **Sobre o MATOPIBA**. Disponível em: < <https://www.embrapa.br/tema-matopiba/sobre-o-tema>>. Acesso em: 30 junho de 2020.

ENEL. **Companhia Energética do Estado do Ceará – COELCE**. Disponível em: <https://www.enel.com.br/pt-ceara.html>. Acesso em: 09 de agosto de 2020.

FERREIRA, M. D. P.; COELHO, A. B. Desmatamento recente nos estados da Amazônia legal: uma análise da contribuição dos preços agrícolas e das políticas governamentais. **Revista de Economia e Sociologia Rural**, v. 53, n. 1, p. 93-108, 2015.

FUNCEME. **Portal Hidrológico do Ceará**. Disponível em: <http://www.hidro.ce.gov.br/municipios/chuvas-diarias>. Acesso em: 09 de agosto de 2020.

FUNDAÇÃO SOS MATA ATLÂNTICA; INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Atlas dos remanescentes florestais da Mata Atlântica: período 2018 – 2019**. São Paulo. 61 p., 2020.

FUNDAÇÃO SOS MATA ATLÂNTICA; INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Aqui tem Mata?** Disponível em: <https://aquitemmata.org.br/#/>. Acesso em: 09 de agosto de 2020.

GOURIEROUX, C., & MONFORT. Simulation-based inference: A survey with special reference to panel data models. **Journal of Econometrics**, v. 59, n. 1-2, p. 5-33, 1993.

GROSSMAN, G.; KRUEGER, A. (1995). Economic growth and the environment. **Quarterly Journal of Economics**, v. 110, n. 2, p. 353-377.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (2017a). **Produto Interno Bruto dos Municípios**. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html?t=downloads>. Acesso em: 09 ago. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (2017b). **Sistema IBGE de Recuperação Automática – SIDRA**. Disponível em: <https://sidra.ibge.gov.br/tabela/5938>. Acesso em 09 ago. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (2017c). **Produção da Extração Vegetal e da Silvicultura – PEVS**. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9105-producao-da-extracao-vegetal-e-da-silvicultura.html?=&t=o-que-e>. Acesso em 09 ago. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (2017d). **Estimativas da População**. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-popul?=&t=downloads>. Acesso em: 09 ago. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (2020). **Pesquisa da Pecuária Municipal - PPM**. Disponível em:

<https://www.ibge.gov.br/estatisticas/economicas/9107-producao-da-pecuaria-municipal.html?=&t=downloads>. Acesso em: 09 ago. 2020.

MINISTÉRIO DA INTEGRAÇÃO. (2018). **Nova delimitação Semiárido**. Disponível em: http://www.sudene.gov.br/images/arquivos/semiariado/arquivos/Relação_de_Municípios_Semiárido.pdf. Acesso em: 09 ago. 2020.

MINISTÉRIO DO MEIO AMBIENTE. (2020). **Painel Unidades de Conservação**. Disponível em: <https://app.powerbi.com/view?r=eyJrIjoiMDNmZTA5Y2ItNmFkMy00Njk2LWI4YjYtZDJINzFkOGM5NWQ4IiwidCI6IjJmY2ZmE5LTNmOTMtNGJiMS05ODMwLTZmNDY3NTJmMDNlNCIsImMiOiJF9>. Acesso em: 09 ago. 2020.

MINISTÉRIO DO TRABALHO. (2017). **A Relação Anual de Informações Sociais – RAIS**. Disponível em: <http://www.rais.gov.br/sitio/index.jsf>. Acesso em: 09 ago. 2020.

OLIVEIRA, R. C.; ALMEIDA, E.; FREGUCLIA, R. S.; BARRETO, R. C. S. Desmatamento e crescimento econômico no Brasil: uma análise da curva de Kuznets ambiental para a Amazônia Legal. **Revista de Economia e Sociologia Rural**, v. 49, n. 3, p. 709-739, 2011.

RODRIGUES, L. A.; CUNHA, D. A.; BRITO, L. M.; PIRES, M. V. Pobreza, crescimento econômico e degradação ambiental no meio urbano brasileiro. **Revista Iberoamericana de Economía Ecológica**, v. 26, n. 1, p. 11-24, 2016.

SISTEMA DE INFORMAÇÕES CONTÁBEIS E FISCAIS DO SETOR PÚBLICO BRASILEIRO. **Contas Anuais**. Disponível em: <http://www.tesouro.fazenda.gov.br/contas-aneais>. Acesso em: 09 de agosto de 2020.

SOARES, L. R., & ALMEIDA, L. T. Desacoplamento de impactos ambientais no Brasil. **Revista Iberoamericana de Economía Ecológica**, v. 28, n. 2, p. 21-43, 2018.

TEIXEIRA, R. F. A. P, BERTELLA, M. A., & ALMEIDA, L. T. Curva de Kuznets ambiental para o estado do mato grosso. **Análise Econômica**, v. 30, n. 57, p. 313-337, 2012.

TOBIN, J. Estimation of relationships for limited dependent variables. **Econometrica**, v. 26, n. 1, p. 24-36, 1958.

APÊNDICE A

Matriz de correlação entre as variáveis explicativas.

	PRECI	RTpc	ICMS pc	FPMp c	CEpc	Emp.p rimário	BOV	DPOP	VPEpc	VAB ^{agro} pc	PIBpc
0,150 *	0,842 *		0,806 *	0,073	0,715 *	- 0,005	- 0,007	0,268 *	- 0,136	0,050	1,000
0,031	- 0,030		- 0,014	0,255 *	0,023	0,445 *	- 0,099	- 0,176	- 0,133	1,000	VAB ^{agro} pc
-	-		-	0,013	-	-	-	-	1,000		VPEpc
0,218	0,122		0,077		0,097	0,118	0,037	0,111			
0,197 *	0,270 *		0,174 *	- 0,165	0,220 *	- 0,084	- 0,056	1,000			DPOP
0,041	-		-	-	-	0,033	1,000				BOV
0,090	0,509		0,571	- 0,043	0,10* 1,000	1,000					Emp.p rimário
0,090	0,509 *		0,571 *	- 0,043	1,000						CEpc
- 0,146	0,120 *		0,427 *	1,000							FPMp c
0,077	0,729 *		1,000								ICMS pc
0,159 *	1,000										RTpc
1,000											PRECI

Fonte: Elaborada com base nos dados da pesquisa.

Nota: * denota a significância ao nível de 5%.

ESCOLARIDADE DOS PAIS E PERFORMANCE ESCOLAR DOS FILHOS AO FINAL DO ENSINO MÉDIO NO CEARÁ

Janaina Rodrigues Feijó*

Valdemar Rodrigues de Pinho Neto**

RESUMO

Esse artigo investiga a relação entre escolaridade dos pais (com ou sem ensino superior) e o desempenho educacional dos filhos ao final do Ensino Médio no Ceará. Em particular, analisamos quatro canais (tamanho da família, renda, infraestrutura domiciliar e escolha das escolas) que intermediam esta relação e mensuramos os efeitos “brutos” e “líquidos” da escolaridade do pai e da mãe sobre a performance dos filhos. Para tanto, utiliza-se microdados do Censo Escolar e do Enem para o ano de 2019, disponibilizados pelo INEP, juntamente com um modelo econométrico que inclui quatro grupos (um para cada possível canal) de efeitos fixos. Os resultados revelam efeitos (brutos) relevantes da educação do pai e da mãe sobre a educação dos filhos e, mesmo após controlar simultaneamente por renda, tamanho da família, infraestrutura domiciliar e escola, a magnitude da influência (líquida) da escolaridade dos pais permanece significativa, embora diminua consideravelmente. Por fim, os efeitos positivos advindos de pais e mães possuírem ensino superior foram substancialmente maiores para a prova de redação e, dentre os possíveis canais, mostramos evidências de que o principal intermediador desse efeito seria as escolas nos quais os filhos são matriculados.

Palavras-Chave: educação dos pais, desempenho dos filhos, Enem.

ABSTRACT

This paper investigates the association between parents' education (with or without higher education) and their children's educational performance at the end of high school in Ceará. In particular, we analyze four channels (family size, income, household infrastructure, and choice of schools) that mediate this relationship and measure the “gross” and “net” effects of father's and mother's schooling on children's performance. To do so, we use microdata from the School Census and Enem for the year 2019, provided by INEP, together with an econometric model that includes four sets (one for each potential channel) of fixed effects. The results show relevant (gross) effects of father's and mother's schooling on the education of their children and, even after controlling simultaneously for income, family size, home infrastructure, and school, the magnitude of the (net) impact of parents' education remains relevant, although it decreases considerably. Ultimately, the positive effects of fathers and mothers holding higher education diplomas were substantially higher for the Essay Test and, among the possible channels, we show evidence that the main mediator of this effect is the schools in which their children are enrolled.

Keywords: Parental education, Educational Performance, Enem.

JEL: I21, I24, I25.

ÁREA 2 - Economia Social.

* Doutora em Economia pelo Programa de Pós-graduação em Economia da Universidade Federal do Ceará (CAEN/UFC).

E-mail: janaina.feijor@gmail.com. Tel.: +55 21 99485 8170

** Doutor em Economia pela Escola Brasileira de Economia e Finanças da Fundação Getúlio Vargas (EPGE/ FGV). Email: valdemar.pinhoneto@gmail.com. Tel.: +55 21 99491 5020.

1 INTRODUÇÃO

A relação entre educação dos pais e desempenho educacional dos filhos tem despertado grande interesse de pesquisadores há pelo menos seis décadas. Os pais desempenham um papel fundamental na vida das crianças, sendo os principais agentes capazes de influenciar a vida educacional e as oportunidades dos filhos (ALMOND E CURRIE, 2010; OREOPOULOS E SALVANES, 2010; CUNHA E HECKMAN, 2007). Esse tema está intrinsecamente relacionado a diversas outras questões sociais. Por exemplo, pais mais escolarizados conseguem influenciar o rendimento escolar dos filhos por meio de inúmeros fatores (ENTWISLE, 2018; BJÖRKLUND E SALVANES, 2011; MO E SINGH, 2008) que tendem a exibir persistência intertemporal e intergeracional, reforçando as desigualdades educacionais e, conseqüentemente, as disparidades de renda (BECKER ET AL., 2018; WITTEVEEN E ATTEWELL, 2017; YANG E QIU, 2016; CORAK, PIRAINO E FERREIRA, 2016; REIS E RAMOS, 2011).

A quantidade de estudos brasileiros analisando essa relação vem crescendo consideravelmente nas últimas décadas, impulsionado pela disponibilidade de novas bases de dados e pelo interesse em compreender mais profundamente os determinantes da performance escolar dos estudantes (CURI E MENEZES FILHO, 2013; FIGUEIRÊDO, NOGUEIRAY E SANTANAZ, 2014, PALERMO, SILVA E NOVELLINO, 2014; SAMPAIO, SAMPAIO, MELLO E MELO, 2011).

A educação dos pais tem sido considerada um fator importante para explicar o desempenho educacional dos indivíduos, mas a literatura tem apontado que há um conjunto de fatores socioeconômicos, em parte correlacionados com a escolaridade dos pais, que se mostram fundamentais no entendimento dessa relação (BJÖRKLUND E SALVANES, 2011; BREDTMANN E SMITH, 2018; GLICK E SAHN, 2000). Nesse sentido, é possível identificar pelo menos quatro canais (observáveis) pelos quais os pais mais educados conseguiriam influenciar (indiretamente) o desempenho educacional dos filhos.

O primeiro deles diz respeito aos recursos financeiros dos pais. Pais mais instruídos tendem a possuir maiores níveis de renda e os trabalhos mostram que há uma relação positiva entre renda familiar e rendimento escolar (BJÖRKLUND E SALVANES, 2011; BREDTMANN E SMITH, 2018; DAVIS-KEAN, 2005; GLICK E SAHN, 2000). O segundo canal está relacionado ao tamanho da família. Considerando a limitação de recursos (de tempo, financeiros etc.) dos pais, o investimento por criança dentro da família diminui à medida que a quantidade de filhos aumenta (CHEN, CHEN E LIU, 2019, BLACK, DEVEREUX, E SALVANES, 2005; LAFORTUNE E LEE, 2014). Dessa forma, espera-se uma relação inversa entre o número de filhos e desempenho educacional das crianças, pois os pais mais escolarizados geralmente apresentam menor número de filhos e, portanto, conseguem investir um maior nível de recursos por filho.

O terceiro canal diz respeito a infraestrutura do domicílio (Curi e Menezes, 2013), que melhora com os níveis de renda dos pais. Por fim, o quarto seria a qualidade das escolas. Pais mais instruídos possuem mais recursos e maior acesso as informações sobre a qualidade das escolas, conseguindo escolher escolas com bons quadros docentes, boa infraestrutura física, melhor gestão etc. Assim, os pais também contribuem para o desempenho educacional das crianças ao escolherem as escolas nas quase eles as matriculam (WOSSMANN, 2016; CURI E MENEZES FILHO, 2013; SCORZAFAVE E FERREIRA, 2011). Mesmo controlando por esses quatro canais observáveis, a educação dos pais ainda mostra alguma influência sobre a performance escolar dos filhos, potencialmente devido a transmissão de um conjunto de valores, referências, expectativas e cultura (BENNER, BOYLER E SANDLER, 2016; WANG, DENG E YANG, 2016; BJÖRKLUND E SALVANES, 2011; JACOBS E HARVEY, 2005), um maior envolvimento parental na educação (BOONK ET AL., 2018; POVEY ET AL., 2016;

CASTRO ET AL, 2015), ou mesmo a carga genética (PLUG, 2004; QIN, WANG E ZHUANG, 2016), dentre outros possíveis (fatores não observados em nossos dados).

Esse trabalho busca mensurar os efeitos das combinações de escolaridade do pai e da mãe na performance dos filhos ao final do ensino médio no Ceará. Para analisar o desempenho dos estudantes cearenses ao final do ensino médio foram utilizados os microdados do Exame Nacional do Ensino Médio (Enem) do ano de 2019, disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

A estratégia empírica consiste em estimar regressões múltiplas com efeitos fixos, visando verificar o padrão de mudanças nos parâmetros que medem o efeito da educação dos pais sobre educação dos filhos à medida que se incluem, isoladamente e conjuntamente, quatro grupos de efeitos fixos (renda familiar, tamanho da família, infraestrutura domiciliar e escola). Quando se inclui todos os efeitos fixos conjuntamente, os parâmetros associados a educação dos pais podem ser entendidos como o efeito “líquido” (sem intermediários) da educação dos pais sobre a educação dos filhos. Assim, compara-se indivíduos semelhantes em termos de renda, tamanho da família, infraestrutura domiciliar e dentro da mesma escola, mas que diferem quanto à escolaridade dos pais. Por outro lado, o “efeito bruto” é medido a partir de um modelo de regressão simples sem a inclusão dos conjuntos de efeitos fixos.

Essa pesquisa busca aprofundar o debate sobre a relação existente entre escolaridade dos pais e desempenho acadêmico dos filhos. Contribuiremos principalmente para a expansão de pesquisas que analisam os estudantes em estágios mais avançados da vida escolar. Os trabalhos averiguando a relevância da educação dos pais geralmente focam em crianças, muito embora haja muitas evidências de que os pais podem influenciar todas as fases da vida escolar dos filhos (BENNER, BOYLE E SADLER, 2016).

De acordo com os resultados encontrados, os filhos de pais e mães com nível superior apresentaram melhor desempenho no Enem, comparado aos filhos de pai e mãe sem nível superior, principalmente na prova de Redação, na qual obtiveram notas em média 80,99% maiores. Embora em magnitude inferior, esse diferencial também foi significativo para as demais provas: Linguagens e Códigos (16,06%), Matemática (27,40%) e Objetivas (20,96%). Após controlar pelos quatro grupos de efeitos fixos (*proxies* para os potenciais canais), a magnitude da influência da escolaridade do pai e da mãe diminuem consideravelmente, e ainda mais acentuadamente na prova de Redação, na qual o efeito líquido dos pais possuem ensino superior passou a ser de apenas 10,32%. O mesmo padrão de queda dos efeitos estimados após controlar pelos intermediadores (tamanho da família, renda, domicílio e escola) observa-se para todas as áreas de conhecimentos da prova, e nota-se a predominância do fator escola em explicar os efeitos da educação dos pais sobre o desempenho dos filhos. Nota-se ainda que as mulheres apresentaram melhor desempenho do que os homens nas provas de Linguagens e Códigos e Redação, e o oposto ocorre em Matemática.

Esse artigo está estruturado em cinco seções, incluindo esta introdução. Na segunda seção tem-se a revisão de literatura e na terceira descreve-se os dados que serão utilizados na análise. Na quarta seção detalha-se a estratégia empírica e na quinta são apresentados os resultados das estimativas. Por fim, na sexta seção tem-se as considerações finais.

2 REVISÃO DE LITERATURA

As pesquisas empíricas vêm mostrando que a performance educacional dos indivíduos está fortemente relacionada com a educação e condições socioeconômicas dos pais. Essa relação é frequentemente verificada nas sociedades para as quais existem dados disponíveis. As contribuições pioneiras de Coleman (1966), em Sociologia, e Becker (1964), na Economia, abriram caminho para uma gama de outros estudos em vários campos da ciência (WILLIAMS, 1980; EPSTEIN, 1987; TEACHMAN, 1987; USEEM, 1992). Dada a sua abrangência e interdisciplinaridade, nota-se o quão complexo é analisar a relação entre nível educacional dos pais e desempenho acadêmico dos filhos.

Os pais mais educados conseguem estruturar melhor a vida educacional dos filhos, uma vez que eles são os principais atores que combinam seus recursos e investem na capacidade de ganhos futuros dos filhos por meio do nível educacional (BREDTMANN E SMITH, 2018; BJÖRKLUND E SALVANES, 2011; OREOPOULOS E SALVANES, 2010). Além disso, pais mais educados passam mais tempo cuidando dos seus filhos, mesmo estando empregados (ENGLAND E SRIVASTAVA, 2013). Ainda que a literatura mostre consenso da importância do nível educacional do pai e da mãe no sucesso escolar dos filhos, ainda não está claro qual a influência relativa de cada um. Na maior parte do debate envolvendo esse tema (KALMIJN, 1994; CROOK, 1995; BJÖRKLUND E RICHARDSON, 2001; KORRUP, GANZEBOOM E VAN DER LIPPE, 2002; BEHRMAN E ROSENZWEI, 2002; MARKS, 2008), os resultados nem sempre mostram um consenso. Geralmente isso ocorre devido às diferentes definições de *background* familiar¹, amostra e contextos culturais entre as pesquisas.

Por exemplo, enquanto Kalmijin (1994) concluiu que a educação da mãe é tão importante quanto a educação do pai nos Estados Unidos, Crook (1995) encontrou efeitos mais expressivos para a educação da mãe na Austrália. Os resultados de Plug (2004) apoiam a ideia de que a influência positiva da escolaridade da mãe sobre a de seu filho desaparece quando habilidades hereditárias e *assortative mating*² são levadas em consideração. Já Marks (2008) mostrou que o impacto da educação da mãe é geralmente maior ou comparável à educação do pai e que a importância relativa das características da mãe aumentou com o tempo.

Apesar de o nível educacional dos pais ser considerado o fator mais importante na determinação dos resultados educacionais das crianças (RUMBERGER, 1995; BJÖRKLUND E SALVANES, 2011; MARBUAH, 2016; MENDES E KARRUZ, 2016), a literatura também aponta que a magnitude desse impacto é grande, e muitas vezes superestimada, porque a influência dos pais ocorre via canais indiretos fortemente correlacionados com seus níveis educacionais. Nesse sentido, os resultados de Santos, Mariano e Costa (2018) concluem que, devido aos retornos privados do capital humano, pais mais escolarizados transmitem para seus filhos uma parcela relevante das vantagens educacionais por meio do nível socioeconômico familiar.

¹ A escolaridade dos pais é vista como um dos principais componentes do *background* socioeconômico familiar. Há várias abordagens distintas para sua medição. Para Barros *et alii* (2006) e Reis e Ramos (2011) a definição engloba renda familiar e escolaridade dos pais. Já Sousa, Oliveira e Annegues (2018) e Marks (2008) adiciona a estas a ocupação dos pais. Figueirêdo, Nogueira e Santana (2014) consideram apenas escolaridade e ocupação dos pais. Björklund e Salvanes, (2011) o define como um amplo conjunto de fatores que o indivíduo não escolheu para si mesmo e, portanto, não pode ser responsabilizado em um sentido normativo.

² O *assortative mating* diz respeito às evidências empíricas de que os indivíduos tendem a escolher parceiros com características semelhantes às suas.

2.1 Canais de transmissão da escolaridade dos pais: fatores observáveis

Com base nos estudos relacionados a esse tema, é possível identificar pelo menos quatro canais pelos quais a escolaridade dos pais pode afetar a educação dos filhos. Pais mais educados podem contribuir na educação dos filhos por meio da: a) renda, b) quantidade de filhos, c) infraestrutura domiciliar e d) escolha da escola.

Renda

A diferença de desempenho entre as crianças pode estar relacionada a disponibilidade de recursos financeiros das suas famílias (GLICK E SAHN, 2000; BJÖRKLUND E SALVANES, 2011; NOBLE ET AL., 2015; BREDTMANN E SMITH, 2018; YANG E QIU, 2016; MARBUAH, 2016). O tempo em que o investimento é realizado também pode potencializar o efeito dessa correlação. Cunha e Heckman (2007) observam a presença de períodos críticos para investimento no desenvolvimento de certas habilidades e que é possível encontrar heterogeneidade no efeito da renda ao longo da vida da criança.

No Brasil, autores também verificaram uma relação positiva entre renda familiar e desempenho educacional (BARROS ET AL., 2006; REIS E RAMOS, 2011; MELO E ARAKAWA, 2016; BARBOSA E SOUSA, 2014; MENDES E KARRUZ, 2016). Barros et al. (2006) ressaltam que a escolaridade dos pais está mais relacionada à renda permanente da família do que à renda *per capita* corrente e a educação é resultado de investimentos acumulados ao longo dos anos. Além disso, o efeito renda pode variar ao longo da distribuição das notas (Barbosa e Souza, 2014). Vale salientar que a renda tem um papel importante no processo de transmissão das desigualdades de rendimentos entre gerações, onde os trabalhadores cujos pais ou mães alcançaram níveis mais altos de educação tendem a apresentar não apenas mais anos de estudo, em média, como também maiores retornos à escolaridade (REIS E RAMOS, 2011).

Tamanho da família

Estudos também vêm apontado que o tamanho da família pode exercer influência sobre os anos de escolaridade das crianças (BLACK, DEVEREUX, E SALVANES, 2005; BJÖRKLUND E SALVANES, 2011; LAFORTUNE E LEE, 2014; CHEN, CHEN E LIU, 2019). Geralmente, as funções de produção da qualidade infantil levam em consideração o insumo tamanho da família. Há um *trade-off* entre quantidade e qualidade infantil dentro de uma família, onde um menor tamanho do agregado familiar está associado a uma média maior de anos de escolaridade. E essa correlação permanece quando se controla por fatores socioeconômicos. Em termos de investimentos, os pais podem investir de forma diferente em seus filhos (diferenças de gêneros /ordem de nascimento), bem como em quantos filhos eles decidem ter, trocando quantidade por qualidade. Além disso, em uma família com menos filhos, o nível de investimento em educação por criança tende a ser maior. Vale notar que o tamanho da família pode ser endógeno e relacionado com outras características parentais não observadas que afetam os *outcomes* das crianças.

Infraestrutura Domiciliar

Os pais mais escolarizados e com maiores níveis de renda podem contribuir indiretamente para a performance do seu filho ao disponibilizar uma melhor infraestrutura física domiciliar que facilita e estimula o aprendizado, como espaços para o estudo, acesso à internet, computadores, dentre outros equipamentos. No Brasil, as pesquisas têm incorporado em seus modelos econométricos algumas dessas características (SOUZA, OLIVEIRA E ANNEGUES, 2018; CURI E MENEZES-FILHO, 2013; PALERMO, SILVA E NOVELINO, 2014). Os resultados de Menezes-Filho (2007) corroboram tal correlação positiva e mostram que, especificamente, alguns itens relacionados ao domicílio ajudaram a explicar o desempenho escolar, entre eles o número de livros e a presença de computador em casa.

Qualidade da Escola

Outro canal pelo qual os pais mais educados conseguem contribuir para o sucesso acadêmico do filho é por meio das escolas. Pais mais educados tendem a possuir tanto disponibilidade financeira para pagar as melhores escolas como também conseguem mais facilmente acesso à informação sobre as características das escolas existentes, investindo mais e melhor. Jacobs e Harvey (2005) verificaram que os grupos de pais das escolas de alto desempenho estavam totalmente cientes dos resultados das escolas de seus filhos nos últimos anos, conseguiam descrever os resultados com precisão e relatavam suas opiniões sobre o porquê de os resultados de suas escolas serem altos. Em contraste, muitos grupos de pais de escolas de médio e baixo desempenho afirmaram que não sabiam quais eram os resultados.

Em relação ao corpo docente das escolas, embora Abdul-Hamid (2007) conclua que a certificação dos professores não tem um impacto estatisticamente significativo nos resultados da aprendizagem, De Hoyos, Espino e García (2012) mostram que a proporção de professores e diretores escolares com pós-graduação estava associada positivamente com os resultados dos testes em matemática. Já Marshal e Sorto (2012) verificaram, para Guatemala e Peru, a importância do domínio do conteúdo dos professores no processo de aprendizagem dos alunos. Glewwe et al. (2014) averiguaram que, dentre os estudos que estimaram o efeito da experiência do professor na aprendizagem dos alunos, 20 mostraram impactos estatisticamente significativos, sendo 17 positivos. Ao analisar as diferenças de desempenho escolar entre um grupo de países, a pesquisa de Woessmann (2016) sugere que as diferenças nos gastos e no tamanho da turma têm um papel limitado na explicação das diferenças de performance entre os países e que as diferenças na qualidade do professor e no tempo de instrução são importantes. Além disso, os alunos se saem pior em escolas onde a capacidade de fornecer instrução é prejudicada pela escassez ou inadequação de materiais instrucionais, como livros didáticos.

No Brasil, Menezes-Filho (2007) encontra que a escola explica entre 10% e 30% das diferenças das notas obtidas pelos alunos e que as escolas privadas tendem a se destacar em relação às instalações físicas (infraestrutura e equipamentos), corpo docente e gestão. Embora as variáveis no âmbito da escola muitas vezes não expliquem grande parte do resultado dos testes (Rivkin, Hanushek e Kain, 2005, Nieto e Ramos, 2014), elas ainda devem ser consideradas importantes. Há grandes diferenças entre as escolas públicas e privadas e entre as escolas localizadas no meio rural e urbano que explicam as diferenças de resultados entre os estudantes. Os resultados de Mendes e Karruz (2016) e Melo e Arakawa (2016) apontam na mesma direção, revelando que quando se amplia a definição de *background* familiar, a rede de ensino é o fator que possui o impacto mais expressivo sobre o desempenho.

2.2. Canais de transmissão da escolaridade dos pais: fatores não observáveis

Diante do que foi discutido até aqui, nota-se que a escolaridade dos pais tende a influenciar o desempenho de seus filhos. Quando não se considera os canais subjacentes a essa relação (como renda, tamanho da família, infraestrutura domiciliar e escolha das escolas), tem-se o efeito “bruto” da educação dos pais. Porém, será que ao controlar pelo máximo de variáveis relacionadas a esses quatro grupos, a educação dos pais ainda conseguiria influenciar a performance educacional dos filhos? Nos referimos a esse efeito (após controlar por esses canais) como o efeito “líquido” da educação dos pais sobre as notas dos filhos.

A literatura vem apontando, principalmente no campo da sociologia e psicologia, que pais mais educados apresentam um conjunto de fatores, muitas vezes subjetivos e de difícil mensuração, que são imprescindíveis para o sucesso dos filhos. No presente estudo denomina-se essa parte como efeito líquido da educação dos pais. Björklund e Salvanes (2011) ressaltam que as origens culturais das famílias, incluindo fatores não observados, como preferências de

risco, preferências de tempo e habilidades parentais, podem afetar as oportunidades das crianças. Os pais também podem transmitir habilidades cognitivas e genéticas não observadas (PLUG, 2004; QIN, WANG E ZHUANG, 2016; SCHEEREN, DAS E LIEFBROER, 2017).

Outro fator destacado pelos pesquisadores como extremamente relevante é o envolvimento dos pais (BOONK ET AL., 2018; POVEY ET AL., 2016; CASTRO ET AL., 2015; JEYNES, 2015). Os pais conseguem melhorar o desempenho acadêmico dos seus filhos quando se envolvem na sua aprendizagem, colaborando com a escola e realizando atividades que incentivam o desenvolvimento das capacidades cognitivas da criança desde seu nascimento. Além disso, envolvem-se em suas experiências e ambientes escolares (POVEY ET AL., 2016; CASTRO ET AL., 2015). Essa colaboração entre pais e professores sinaliza para as crianças a importância da educação e propicia uma continuidade de ações entre casa e escola (EPSTEIN E LEE, 1995; EPSTEIN, 1987; SCOTT-JONES, 1995). Wang e Sheikh-Khalil (2014) verificam que o envolvimento dos pais explicou o sucesso acadêmico do adolescente e a saúde mental direta e indiretamente por meio de engajamento comportamental e emocional.

As atitudes e as expectativas dos pais também contribuem para a performance dos filhos (BENNER, BOYLE E SADLER, 2016; WAN, DENG E YANG, 2016; CASTRO ET AL., 2015; WILDER, 2014). De acordo com Wilder (2014), as expectativas dos pais refletem suas crenças e atitudes para com a escola, professores, disciplinas e educação em geral. Como as crianças tendem a ter atitudes e crenças semelhantes às dos pais, ter expectativas parentais elevadas é fundamental para o desempenho acadêmico das crianças. Corroborando tal resultado, Castro et al. (2015) acrescentam que as associações mais fortes também são encontradas quando as famílias desenvolvem e mantêm a comunicação sobre suas atividades escolares e as ajudam a desenvolver hábitos de leitura. Wang, Deng e Yang (2016) salientam que as expectativas dos pais conseguem, muitas vezes, mediar a relação entre o *status* econômico da família e o envolvimento dos pais.

Jacobs e Harvey (2005) identificaram que o alto desempenho dos estudantes foi explicado principalmente pelas expectativas dos pais sobre o nível de escolaridade dos filhos, seguida pelo tempo em que eles mantiveram essas expectativas, que geralmente iniciava no nascimento e permanecia até eles frequentarem a universidade. Além disso, os autores constataram que todos os pais dos alunos das escolas de maior desempenho tinham frequentado a universidade, enquanto a maioria dos pais (70,9%) na escola de menor rendimento alcançaram apenas o nível do ensino secundário. Os resultados do estudo sugerem que os alunos academicamente bem-sucedidos provavelmente vêm de ambientes familiares em que seus pais têm um histórico acadêmico forte e provavelmente têm altas aspirações acadêmicas e profissionais para seus filhos.

3 DADOS E ESTRATÉGIA EMPÍRICA

3.1 Dados

Para o desenvolvimento desse estudo foi utilizado os microdados do Exame Enem 2019. Esse exame busca averiguar se os participantes apresentam domínio dos princípios científicos e tecnológicos que guiam a produção moderna e se possuem conhecimento das formas contemporâneas de linguagem. O Exame é composto por quatro provas objetivas (Linguagens e Códigos, Matemática, Ciências da Natureza e Ciências Humanas) e uma redação. Cada uma das provas objetivas contém 45 questões de múltipla escolha. Mais detalhes sobre o Exame Enem podem ser encontrados no Anexo A.

A amostra de interesse foi composta inicialmente por 112.068 estudantes residentes do Ceará que estavam finalizando o ensino médio em alguma escola do Ceará no ano de 2019.

Contudo nem todos realizaram todas as provas. Cerca de 91.928 (82%) dos estudantes fizeram as provas de Linguagens e Códigos, Ciências Humanas e Redação, e 88.555 (79%) estavam presentes nas provas de Matemática e Ciências da Natureza. Sendo que 88.080 destes fizeram as cinco provas.

Para que seja possível analisar os efeitos da escolaridade do pai e da mãe conjuntamente no desempenho educacional dos filhos é necessário que todos os participantes tenham identificado, ao preencher o questionário socioeconômico no ato da inscrição, tanto o nível educacional do pai quanto o da mãe. Sendo assim, considerando apenas aqueles que sabiam a escolaridade dos pais, a amostra final a ser analisada é constituída de 73.086 estudantes que estavam presentes nas provas de Linguagens e Código, Ciências Humanas e Redação e 70.613 nas provas de Matemática e Ciências da Natureza. Por fim, 70.255 realizaram as cinco provas. Optou-se por não excluir da amostra os estudantes que não realizaram todas as cinco provas, pois isso poderia incorrer em viés de seleção, já que aumentaria a probabilidade de selecionar os candidatos com melhores desempenhos. O Quadro 1 apresenta a descrição das variáveis.

Quadro 1- Descrição das variáveis

Ti po	Variável	Descrição	
Variáveis	LC	Logaritmo natural da nota da prova de Linguagens e Códigos	
	MT	Logaritmo natural da nota da prova de Matemática	
	OBJ	Logaritmo natural da nota em Objetivas (média simples das provas Linguagens e Códigos, Matemática, Ciências da Natureza e Ciências Humanas).	
	RED	Logaritmo natural da nota da prova de Redação	
	CH	Logaritmo natural da nota da prova Ciências e Tecnologias	
	CN	Logaritmo natural da nota da prova de Ciências Humanas	
Grupos de Controles	Mulher	Variável binária. Assume valor 1 se o estudante for do gênero feminino e "0" caso contrário.	
	Negro_pardo	Variável binária. Assume valor 1 se o estudante se autodeclara negro ou pardo e "0" caso contrário.	
	Renda	Representa a renda familiar mensal do participante. É classificada em 17 categorias ¹	
	Tamanho da família	Representa a quantidade de pessoas que moram na residência do estudante. A quantidade varia de 1 (participante mora sozinho) até 20.	
	Infraestrutura do domicílio	Banheiro	"1" se tiver mais de 1 banheiro na residência e "0" caso contrário
		Quarto	"1" se tiver mais de 2 quartos na residência e "0" caso contrário
		Carro	"1" se tiver carro(s) e "0" caso contrário
		Máquina de Lavar Roupas	"1" se tiver máquina de lavar roupa e "0" caso contrário
		Micro-ondas	"1" se tiver micro-ondas e "0" caso contrário
		Aspirador de Pó	"1" se tiver aspirador(es) de pó e "0" caso contrário
		Televisão em cores	"1" se tiver mais de 1 uma tv na residência e "0" caso contrário
		Aparelho de DVD	"1" se tiver aparelho(s) de DVD e "0" caso contrário
		Tv por assinatura	"1" se tiver TV por assinatura e "0" caso contrário
		Telefone Celular	Quantidade de telefone(s) celulare(s) na residência
		Telefone fixo	"1" se tiver telefone(S) fixo(s) e "0" caso contrário
Computador	"1" se tiver pelo menos 1 computador e "0" caso contrário		
Acesso à Internet	"1" se tiver acesso à internet e "0" caso contrário		
Escola	Identificador da escola onde o participante estava matriculado.		
Educação	Educ_m	Nível educacional da mãe ou da mulher responsável pelo estudante, que assume valor "1" se ela possui pelo menos graduação e "0" caso contrário.	
	Educ_p	Nível educacional do pai ou homem responsável pelo estudante, que assume valor "1" se ele possui pelo menos graduação e "0" caso contrário.	

Elaboração: própria da autora com base no INEP. Nota: 1) Nenhuma renda, 2) Até R\$ 998,00, 3) R\$ 998,01 até R\$ 1.497,00, 4) R\$ 1.497,01 até R\$ 1.996,00, 5) R\$ 1.996,01 até R\$ 2.495,00, 6) R\$ 2.495,01 até R\$ 2.994,00, 7) R\$ 2.994,01 até R\$ 3.992,00, 8) R\$ 3.992,01 até R\$ 4.990,00, 9) R\$ 4.990,01 até R\$ 5.988,00, 10) R\$ 5.988,01 até R\$ 6.986,00, 11) R\$ 6.986,01 até R\$ 7.984,00, 12) R\$ 7.984,01 até R\$ 8.982,00, 13) R\$ 8.982,01 até R\$ 9.980,00, 14) R\$ 9.980,01 até R\$ 11.976,00, 15) R\$ 11.976,01 até R\$ 14.970,00, 16) R\$ 14.970,01 até R\$ 19.960,00, 17) Mais de R\$ 19.960,00.

3.2 Estratégia Empírica

A estratégia empírica consiste em estimar um modelo de regressão linear com efeitos fixos dado por:

$$Y_{irdfe} = \omega_r + \tau_d + \pi_f + \gamma_e + \beta^m E_i^m (1 - E_i^p) + \beta^p E_i^p (1 - E_i^m) + \beta^{mp} E_i^m E_i^p + \varepsilon_{irdfe} \quad (1)$$

Onde Y_{irdfe} é a pontuação no Enem do indivíduo i de classe de renda r , infraestrutura domiciliar do tipo³ d , tamanho da família f e que estuda na escola e . Neste trabalho, o *outcome* de interesse, Y_{irdfe} , será representado pela pontuação nas provas de Linguagens e Códigos, Ciências Humanas, Matemática, Ciências da Natureza, Redação e Objetivas. O modelo conta com quatro dimensões de efeitos fixos (ω_r , τ_d , π_f e γ_e), que serão utilizados para “isolar” o efeito direto da educação dos pais sobre o desempenho dos filhos nas provas.

As variáveis explicativas de interesse⁴ são *dummies* que assumem valor “1” caso a mãe (E_i^m) ou o pai (E_i^p) do indivíduo i tenha ensino superior completo. Essas variáveis entram no modelo de forma interativa, de modo que $E_i^m (1 - E_i^p)$ indica “apenas a mãe possui ensino superior”, $E_i^p (1 - E_i^m)$ significa que “apenas o pai possui ensino superior” e $E_i^m E_i^p$ significa que ambos possuem ensino superior completo. Dessa forma, torna-se possível entender a importância relativa da educação de cada um dos pais isoladamente. O termo ε_{irdfe} é o erro idiossincrático do modelo, representando todos os fatores que podem afetar as notas dos alunos, mas que não estejam mensuradas diretamente no modelo (nos efeitos fixos ou na educação dos pais).

Dito isto, os principais parâmetros de interesse dessa pesquisa são β^m , β^p e β^{mp} . O coeficiente β^m mede o efeito de apenas a mãe possuir ensino superior sobre o desempenho dos filhos. Analogamente, β^p mede o efeito de apenas o pai ter ensino superior sobre as notas dos filhos. Por fim, β^{mp} mede o efeito de ambos (pai e mãe) possuírem ensino superior completo. Note que, nesse modelo, devido a multicolinearidade, não é possível incluir uma variável indicando o caso em que ambos (pai e mãe) não possuem ensino superior. Portanto, a interpretação dos coeficientes descritos acima será sempre relativa a essa categoria base/omitida, ou seja, relativa ao caso em que os indivíduos são provenientes de famílias nas quais ambos (pai e mãe) não têm ensino superior. Denominou-se efeito líquido quando se controla simultaneamente por esses quatro efeitos fixos. Matematicamente, os efeitos capturados a partir desse modelo são dados por:

Efeito “líquido” de apenas a mãe possuir ensino superior:

$$\beta^m \equiv E[Y_{irdfe} | \omega_r, \tau_d, \pi_f, \gamma_e, E_i^m = 1, E_i^p = 0] - E[Y_{irdfe} | \omega_r, \tau_d, \pi_f, \gamma_e, E_i^m = 0, E_i^p = 0] \quad (2)$$

Efeito “líquido” de apenas o pai possuir ensino superior:

$$\beta^p \equiv E[Y_{irdfe} | \omega_r, \tau_d, \pi_f, \gamma_e, E_i^m = 0, E_i^p = 1] - E[Y_{irdfe} | \omega_r, \tau_d, \pi_f, \gamma_e, E_i^m = 0, E_i^p = 0] \quad (3)$$

³ Para a construção do “tipo de domicílio” definiu-se uma variável que resume a informação de todas as características de infraestrutura domiciliar presentes no Quadro 1. Em suma, considera-se que um domicílio é igual a outro quando eles são iguais em TODAS as dimensões (i.e., possuem exatamente os mesmos itens). Analogamente, domicílios que se distinguem em pelo menos um dos itens, são tratados como domicílios diferentes.

⁴ Por simplicidade de notação, omitiu-se os demais índices (r, d, f, e).

Efeito “líquido” de ambos (pai e mãe) possuírem ensino superior:

$$\beta^{mp} \equiv E[Y_{irafe} | \omega_r, \tau_d, \pi_f, \gamma_e, E_i^m = 1, E_i^p = 1] - E[Y_{irafe} | \omega_r, \tau_d, \pi_f, \gamma_e, E_i^m = 0, E_i^p = 0] \quad (4)$$

Note que nos efeitos capturados acima controla-se por vários canais observáveis (renda, infraestrutura domiciliar, tamanho da família e escola) pelos quais a educação dos pais pode estar influenciando o desempenho dos filhos. Assim, o efeito remanescente pode ser entendido como um efeito mais próximo possível do efeito direto que a educação dos pais tem sobre a educação de seus filhos. Além de mensurar esses efeitos “líquidos”, a ideia deste artigo é verificar o padrão de mudanças observados nos parâmetros β^m , β^p e β^{mp} (que medem o efeito da educação dos pais sobre educação dos filhos) à medida que se incluem (isoladamente e conjuntamente) os quatro grupos de efeitos fixos. Em suma, quando se inclui todos os efeitos fixos conjuntamente os parâmetros β 's podem ser entendidos como o efeito “líquido” da educação dos pais sobre a educação dos filhos, no sentido de não ser um efeito intermediado por nenhuma das quatro características destacadas acima.

Por outro lado, o “efeito bruto” da educação dos pais sobre o desempenho dos filhos pode ser medido a partir de um modelo sem a inclusão dos vários efeitos fixos considerados no modelo da Equação 1.1. Nesse caso temos:

Efeito “bruto” de apenas a mãe possuir ensino superior:

$$\beta^m \equiv E[Y_{irafe} | E_i^m = 1, E_i^p = 0] - E[Y_{irafe} | E_i^m = 0, E_i^p = 0] \quad (5)$$

Efeito “bruto” de apenas o pai possuir ensino superior:

$$\beta^p \equiv E[Y_{irafe} | E_i^m = 0, E_i^p = 1] - E[Y_{irafe} | E_i^m = 0, E_i^p = 0] \quad (6)$$

Efeito “bruto” de ambos (pai e mãe) possuírem ensino superior:

$$\beta^{mp} \equiv E[Y_{irafe} | E_i^m = 1, E_i^p = 1] - E[Y_{irafe} | E_i^m = 0, E_i^p = 0] \quad (7)$$

Comparando os β 's que mensuram os efeitos *brutos* e *líquidos*, como descrito acima, é possível perceber a importância relativa dos quatro componentes/mecanismos na determinação da relação entre educação dos pais e desempenho dos filhos. Nesse artigo, os efeitos fixos são incluídos separadamente, para avaliar a contribuição individual de cada mecanismo, e também conjuntamente, com o objetivo de entender o efeito final após controlar por todos os fatores.

4 RESULTADO E DISCUSSÃO

4.1 Estatística Descritiva

A tabela com a estatística descritiva pode ser encontrada no Anexo B. Verifica-se que as maiores pontuações médias foram obtidas em Redação (541,152) e Matemática (504,787), enquanto que as menores foram observadas em Ciências da Natureza (455,76) e Linguagens e Códigos (482,534). Os maiores desvios padrões também estão associados à performance dos estudantes em Redação e Matemática⁵. Cerca de 8,18% dos participantes tinham mãe com pelo menos ensino superior completo, enquanto 4,99% tinham pai com esse nível educacional.

⁵ Apenas redação possui obrigatoriamente o limite superior igual a 1000 pontos. Isso acontece porque o Enem não trabalha com o percentual de acertos (teoria clássica) e sim com a TRI para a correção das provas objetivas. Nessa teoria, a unidade fundamental de análise é a questão. Com isso, as provas não possuem valores máximos fixos.

Observa-se também que 89,75% dos estudantes eram filhos de pai e mãe que não tinham ensino superior completo. Os que possuíam ambos os responsáveis com pelo menos graduação representavam 2,92% da amostra. Os participantes, em sua maioria, eram do gênero feminino (57,65%), autodeclarados negro ou pardo (78,25%). O tamanho médio das famílias era de quatro pessoas.

No grupo de infraestrutura domiciliar, um pouco mais da metade dos estudantes tinha em suas residências acesso à internet (55,34%). Cerca de 18,70% dos estudantes moravam em casas com 3 ou mais quartos e mais de 17% tinham em casa mais de dois banheiros e possuíam pelo menos um carro em casa. Apenas uma parcela pequena dos estudantes possuía tv por assinatura (10,15%), pelo menos uma máquina de lavar (26,18%), pelo menos um micro-ondas (22,65%) e pelo menos um computador (23,04%). Analisando as classes de renda, nota-se que a maior parte dos estudantes residia em domicílios com renda familiar mensal de R\$0- 998,00 (57,32%) e R\$ 998,01-1.996,50 (27,6%). Ainda sob esse aspecto, verifica-se que cerca de 92,8% dos estudantes estavam inseridos em uma das cinco primeiras classes de renda, ou seja, em residências com renda familiar mensal entre R\$ 0 e R\$ 2.495,00.

4.2 Resultados

As Tabelas 1 a 4 apresentam as estimações das regressões que buscam analisar o efeito da escolaridade do pai e da mãe, conjuntamente, sobre a performance dos seus filhos nas provas do Enem. Os resultados para Ciências Humanas e Ciências da Natureza encontram-se no Anexo C. O modelo segue o padrão log-linear, onde as variáveis dependentes são o logaritmo natural das pontuações⁶ em cada uma das provas. Todas as variáveis explicativas são binárias e as equações foram estimadas por MQO com efeitos fixos. A variável pai e mãe sem ensino superior (*educ_p0_m0*) foi omitida no modelo e a interpretação do impacto das combinações de nível educacional do pai e da mãe será relativo a essa combinação omitida. Todas as tabelas dessa subseção seguem a mesma estrutura, onde na primeira coluna têm-se as estimações sem adicionar nenhum dos quatro grupos de variáveis de controles (efeitos fixos). Nas demais colunas adiciona-se, sequencialmente, cada um dos grupos e na última coluna todos os grupos de efeitos fixos são adicionados simultaneamente. Todos os coeficientes foram estatisticamente diferentes do valor nulo ao nível de significância de 1%.

Na coluna 1 dizemos que o impacto da educação do pai e da mãe seria um efeito “bruto”, pois existem muitos fatores correlacionados com seus níveis educacionais que favorecem o desempenho dos filhos e não estão sendo considerados explicitamente na regressão. Nas colunas 2 a 5, analisa-se como esse impacto é alterado à medida que se adicionam os controles. Os resultados da coluna 6 diz respeito ao efeito líquido, uma vez que se compara alunos que possuem a mesma estrutura domiciliar, renda familiar parecida, mesmo tamanho da família e que estudam na mesma escola, mas com escolaridade do pai e da mãe diferentes. Sendo assim, o impacto da educação do pai e da mãe é o mais “puro” que se consegue chegar ao controlar pelo maior conjunto de variáveis observáveis consideradas relevantes na literatura e disponíveis nos microdados do Enem 2019. Logo, os coeficientes na coluna 6 capturam exatamente o que não é explicado pelos controles, ou seja, o efeito “líquido” da educação dos pais.

Na prova Linguagens e Códigos (Tabela 1), em termos de efeito bruto, verificou-se que os filhos de pai e mãe com pelo menos nível superior completo (*educ_p1_m1*) tinham, em média, desempenho 16,06% maior do que os filhos de pai e mãe sem graduação completa (*educ_p0_m0*). Nota-se também que o impacto na nota de Linguagens e Códigos foi maior para alunos que tinham apenas o pai com ensino superior (10,11%), quando comparado com os que

⁶ Definiu-se como variável dependente o $\ln(1+Y)$, onde Y é a pontuação dos inscritos. Uma vez que o logaritmo natural não é definido para valores iguais a zero, optou-se por essa normalização.

tinham apenas mãe com ensino superior (8,29%). Os estudantes do gênero feminino tiveram, em média, uma pontuação 2% superior as pontuações registradas do gênero masculino. Os autodeclarados negro ou pardo registraram pontuações inferiores as demais raças em 3%.

Tabela 1: Impacto da escolaridade dos pais no desempenho educacional dos filhos. Variável dependente: Linguagens e Códigos. Prova Enem 2019, Ceará.

Variáveis	(1)	(2)	(3)	(4)	(5)	(6)
educ_p0_m1	0,08929*** (0,00339)	0,04745*** (0,00353)	0,08704*** (0,00340)	0,04598*** (0,00372)	0,03346*** (0,00342)	0,01816*** (0,00375)
educ_p1_m0	0,10116*** (0,00532)	0,05110*** (0,00544)	0,09976*** (0,00532)	0,04670*** (0,00579)	0,03223*** (0,00528)	0,01688*** (0,00574)
educ_p1_m1	0,16067*** (0,00450)	0,07178*** (0,00538)	0,15852*** (0,00450)	0,07971*** (0,00551)	0,04240*** (0,00525)	0,02453*** (0,00590)
mulher	0,02131*** (0,00161)	0,02506*** (0,00159)	0,02139*** (0,00161)	0,02378*** (0,00164)	0,01791*** (0,00155)	0,01929*** (0,00161)
negro_pardo	- (0,00194)	- (0,00194)	- (0,00194)	- (0,00203)	- (0,00199)	- (0,00209)
constante	6,19510*** (0,00197)					
observações	73.086	73.086	73.086	71.060	73.034	71.004
R ²	0,03757	0,05909	0,03957	0,10991	0,13572	0,18317
Renda	No	Yes	No	No	No	Yes
Tamanho da Família	No	No	Yes	No	No	Yes
Infraestrutura	No	No	No	Yes	No	Yes
Escola	No	No	No	No	Yes	Yes

Fonte: Elaboração dos autores com base nos microdados do ENEM/INEP 2019.

Notas: ¹Erros padrões em parênteses. *** p<0.01, ** p<0.05, * p<0.1. ² Educ_p0_m1: pai sem nível superior e mãe com nível superior. Educ_p1_m0: pai com nível superior e mãe sem nível superior. Educ_p1_m1: ambos com nível superior.

Controlando apenas por renda (Coluna 2), pode-se observar que a magnitude do impacto cai consideravelmente para todas as três categorias de educação parental. Ou seja, grande parte da contribuição bruta da educação do pai e da mãe está associada ao efeito renda. Esse padrão de comportamento é ainda mais forte quando se controla apenas por efeitos fixos de escolas (Coluna 5), reduzindo o efeito de o pai e a mãe terem nível superior de 16,06% para 4,24%. Já em relação ao controle de tamanho da família (coluna 3), nota-se que os efeitos não diferem substancialmente dos apresentados na coluna 1.

Ao adicionar todos os quatros grupos de controles, observa-se que o efeito do pai e a mãe possuírem nível superior foi de 2,45%. Também ocorreu reduções no efeito da educação do pai e da mãe, tanto para os que possuíam apenas pai com nível superior (10,11% para 1,68%) quanto para os que tinham apenas mãe com nível superior (8,92% para 1,81%). Outro fato interessante é que o diferencial de desempenho entre negros/pardos e as demais raças caiu de -3% para -1%. Mostrando que as desvantagens de performance relacionadas a raça praticamente desaparecem quando se controla pelos quatro efeitos fixos.

O impacto bruto da educação do pai e da mãe foi ainda maior na prova de Matemática (coluna 1 da Tabela 2), onde os filhos com pai e mãe com graduação registraram desempenho 27,40% maior do que os que não tinham nem pai e nem mãe com ensino superior. Para os que tinham apenas mãe ou apenas pai com graduação esse impacto foi de 13,10% e 14,43%, respectivamente. Nota-se que a influência indireta dos pais se deu fortemente via Renda (coluna 2) e Escola (coluna 5).

Tabela 2: Impacto da escolaridade dos pais no desempenho educacional dos filhos. Variável dependente: Matemática. Prova Enem 2019, Ceará.

Variáveis	(1)	(2)	(3)	(4)	(5)	(6)
educ_p0_m1	0,13103*** (0,00320)	0,06648*** (0,00327)	0,12913*** (0,00320)	0,06918*** (0,00347)	0,04220*** (0,00302)	0,02076*** (0,00334)
educ_p1_m0	0,14436*** (0,00501)	0,06333*** (0,00504)	0,14307*** (0,00501)	0,06617*** (0,00540)	0,03405*** (0,00466)	0,01596*** (0,00510)
educ_p1_m1	0,27402*** (0,00423)	0,11533*** (0,00497)	0,27197*** (0,00423)	0,13579*** (0,00513)	0,05657*** (0,00462)	0,02729*** (0,00523)
	-	-	-	-	-	-
mulher	0,04224*** (0,00153)	0,03672*** (0,00149)	0,04215*** (0,00153)	0,03766*** (0,00155)	0,04707*** (0,00138)	0,04380*** (0,00144)
	-	-	-	-	-	-
negro_pardo	0,04291*** (0,00184)	0,02671*** (0,00181)	0,04225*** (0,00184)	0,02187*** (0,00190)	0,01645*** (0,00177)	0,01230*** (0,00186)
constante	6,24019*** (0,00187)					
Observações	70.613	70.613	70.613	68.597	70.560	68.540
R ²	0,10746	0,15854	0,10917	0,20033	0,29821	0,33501
Renda	No	Yes	No	No	No	Yes
Tamanho da Família	No	No	Yes	No	No	Yes
Infraestrutura domiciliar	No	No	No	Yes	No	Yes
Escola	No	No	No	No	Yes	Yes

Fonte: Elaboração dos autores com base nos microdados do ENEM/INEP 2019.

Notas: ¹Erros padrões em parênteses. *** p<0.01, ** p<0.05, * p<0.1. ² Educ_p0_m1: pai sem nível superior e mãe com nível superior. Educ_p1_m0: pai com nível superior e mãe sem nível superior. Educ_p1_m1: ambos com nível superior.

Comparando indivíduos com características de escolas semelhantes⁷, a contribuição da educação do pai e da mãe caiu mais de 67% em relação ao efeito bruto, nas três combinações educacionais. Ou seja, para matemática, a escola em que os pais matriculam seus filhos é relevante para a performance dos filhos. Nota-se que mesmo ao controlar por escolas, os alunos que estudavam em escolas semelhantes, mas com pai e mãe com nível superior reportaram notas 5,65% maiores do que aqueles com pai e mãe sem graduação.

Verifica-se que, mesmo após adicionar todos os controles, filhos de pai e/ou mãe com nível superior tendem a ter melhor desempenho comparado aos filhos de pai e mãe sem esse nível educacional. Além disso, os resultados mostram, como em Linguagens e Códigos, que filhos de pai com graduação, tendem a performar melhor em Matemática do que os filhos de mãe com graduação. Outro ponto a ser destacado é que o desempenho dos homens foi melhor do que o das mulheres, não apresentando diferenças expressivas nos seis modelos estimados. Na tabela 3 tem-se estimativa para Objetivas, que é uma média simples das pontuações nas quatro provas objetivas do ENEM. Embora a magnitude dos coeficientes tenha sido menor do que os encontrados em Matemática, o padrão de resultados permanece.

⁷ Uma vez que foram adicionados efeitos fixos para as escolas, conseguindo diferenciá-las ao máximo possível e comparar alunos dentro da mesma escola.

Tabela 3: Impacto da escolaridade dos pais no desempenho educacional dos filhos. Variável dependente: Objetivas. Prova Enem 2019, Ceará.

Variáveis	(1)	(2)	(3)	(4)	(5)	(6)
educ_p0_m1	0,10385*** (0,00216)	0,05294*** (0,00218)	0,10160*** (0,00216)	0,05342*** (0,00229)	0,03441*** (0,00192)	0,01728*** (0,00211)
educ_p1_m0	0,12090*** (0,00338)	0,05799*** (0,00336)	0,11946*** (0,00338)	0,05758*** (0,00356)	0,03375*** (0,00297)	0,01885*** (0,00322)
educ_p1_m1	0,20958*** (0,00285)	0,09093*** (0,00331)	0,20737*** (0,00285)	0,10392*** (0,00338)	0,04751*** (0,00295)	0,02580*** (0,00330)
mulher	- (0,00103)	- (0,00099)	- (0,00103)	- (0,00102)	- (0,00088)	- (0,00091)
negro_pardo	- (0,00125)	- (0,00121)	- (0,00124)	- (0,00126)	- (0,00113)	- (0,00118)
constante	6,19702*** (0,00127)					
Observações	70.255	70.255	70.255	68.238	70.202	68.181
R ²	0,13158	0,19971	0,13573	0,25325	0,38968	0,43047
Renda	No	Yes	No	No	No	Yes
Tamanho da Família	No	No	Yes	No	No	Yes
Infraestrutura	No	No	No	Yes	No	Yes
Escola	No	No	No	No	Yes	Yes

Fonte: Elaboração dos autores com base nos microdados do ENEM/INEP 2019. Notas: ¹Erros padrões em parênteses. *** p<0.01, ** p<0.05, * p<0.1. ² Educ_p0_m1: pai sem nível superior e mãe com nível superior. Educ_p1_m0: pai com nível superior e mãe sem nível superior. Educ_p1_m1: ambos com nível superior.

De acordo com a Tabela 4, constata-se que o fato de o pai e a mãe terem nível superior eleva a nota de seu filho em 80,99% na prova de Redação, em relação aos que não tem pai e mãe com esse nível educacional. Grande parte da magnitude desse efeito está relacionada a esses pais e mães colocarem seus filhos em boas escolas, pois ao controlarmos por características da escola esse impacto cai para 10,30%. Em Redação as mulheres apresentaram uma pontuação superior aos homens em quase 30%. Em termos de efeito líquido, a magnitude dos impactos da combinação educacional do pai e da mãe teve pequenas diferenças entre si, onde filhos com apenas mãe, apenas pai ou ambos com nível superior apresentaram desempenho 9,31%, 8,31% e 10,30%, respectivamente, maior do que os filhos de pai e mãe sem graduação.

Os resultados apresentados até aqui mostram que, mesmo incorporando no modelo o máximo de características observáveis representando os mecanismos pelos quais os pais podem influenciar indiretamente a performance dos seus filhos, ainda há, impactos significativos da educação parental (efeito líquido). Esse impacto pode estar relacionado a fatores não observáveis que os pais conseguem repassar para seus filhos, como transmissão de valores e expectativas, visão de referência (BENNER ET AL., 2016; WANG ET AL., 2016; JACOBS E HARVEY, 2005; BJÖRKLUND E SALVANES, 2011), cuidados e estímulos às capacidades cognitivas na infância, carga genética dos pais (PLUG, 2004; QIN, WANG E ZHUANG, 2016; SCHEEREN, DAS E LIEFBROER, 2017), dentre outros. Há também uma extensa literatura sobre os benefícios do envolvimento parental (BOONK ET AL., 2018; POVEY ET AL., 2016).

Outro ponto a destacar é que o impacto líquido da educação do pai e da mãe foi maior na prova de Redação enquanto os menores impactos foram observados nas provas de Ciências da Natureza e Ciências Humanas. Esses resultados podem estar relacionados ao fato de que a escola consegue influenciar mais diretamente o desempenho dos alunos em Ciências da Natureza e Ciências Humanas por se tratar de conhecimentos mais específicos. E, assim, ao controlar por escola, o impacto advindo do nível educacional do pai e da mãe é pequeno para essas duas provas. Já em Redação, os pais conseguem transmitir mais conhecimento ao

estimular o desenvolvimento do filho dentro do ambiente familiar por meio de conversas, incentivos à leitura, dentre outros.

Tabela 4: Impacto da escolaridade dos pais no desempenho educacional dos filhos. Variável dependente: Redação. Prova Enem 2019, Ceará.

Variáveis	(1)	(2)	(3)	(4)	(5)	(6)
educ_p0_m1	0,54713*** (0,02705)	0,30681*** (0,02831)	0,53228*** (0,02708)	0,26497*** (0,03007)	0,18656*** (0,02730)	0,09307*** (0,03034)
educ_p1_m0	0,59680*** (0,04242)	0,31809*** (0,04361)	0,58704*** (0,04240)	0,25373*** (0,04683)	0,17414*** (0,04222)	0,08306* (0,04644)
educ_p1_m1	0,80994*** (0,03585)	0,34932*** (0,04312)	0,79498*** (0,03586)	0,34089*** (0,04456)	0,18815*** (0,04198)	0,10302** (0,04777)
mulher	0,28246*** (0,01282)	0,30457*** (0,01277)	0,28239*** (0,01281)	0,30248*** (0,01330)	0,25266*** (0,01238)	0,26457*** (0,01302)
negro_pardo	- (0,01549)	- (0,01554)	- (0,01549)	- (0,01639)	- (0,01594)	- (0,01688)
constante	5,72939*** (0,01572)					
Observações	73.086	73.086	73.086	71.060	73.034	71.004
R ²	0,02336	0,03547	0,02520	0,07833	0,11885	0,15373
Renda	No	Yes	No	No	No	Yes
Tamanho da Família	No	No	Yes	No	No	Yes
Infraestrutura	No	No	No	Yes	No	Yes
Escola	No	No	No	No	Yes	Yes

Fonte: Elaboração dos autores com base nos microdados do ENEM/INEP 2019.

Notas: ¹Erros padrões em parênteses. *** p<0.01, ** p<0.05, * p<0.1. ² Educ_p0_m1: pai sem nível superior e mãe com nível superior. Educ_p1_m0: pai com nível superior e mãe sem nível superior. Educ_p1_m1: ambos com nível superior.

As mulheres tiveram melhor desempenho do que os homens nas provas de Linguagens e Códigos e Redação enquanto os homens foram melhores em Matemática e Ciências da Natureza. Por fim, algumas considerações precisam ser levantadas. Os resultados mostraram que há grandes diferenças quando se analisa a correlação entre escolaridade dos pais e desempenho dos filhos (impacto bruto) e quando se aplica os quatro efeitos fixos simultaneamente (impacto líquido) na performance do Enem. Os dois impactos são relevantes para compreender como o nível educacional dos pais pode afetar o desempenho dos seus filhos. Enquanto o primeiro nos dá uma visão agregada, o segundo nos permite chegar mais perto dos fatores não observáveis pelos quais os pais influenciam a performance dos filhos.

Como teste de robustez, os modelos foram estimados novamente sem excluir os estudantes que responderam não saber a educação dos pais. Considerou-se que os pais desses alunos possuíam escolaridade inferior ao ensino superior. Tal hipótese foi considerada pelo fato de aproximadamente 80% desses estudantes estarem nos quatro primeiros extratos de renda familiar mensal. Considerando essa hipótese, a magnitude dos coeficientes foi um pouco menor do que os resultados apresentados nessa seção, entretanto as conclusões desse artigo permaneceriam substancialmente as mesmas.

5 CONSIDERAÇÕES FINAIS

Esse trabalho teve como objetivo analisar o impacto da escolaridade do pai e da mãe no desempenho dos estudantes cearenses que fizeram o exame Enem em 2019 e que concluiriam o ensino médio nesse mesmo ano. A análise foca nos estudantes ao final da adolescência, ou seja, quando os investimentos dos pais, em termos de transmissão de valores, tempo e recursos financeiros, já estavam mais consolidados. Embora o envolvimento dos pais diminua à medida que os filhos avançam nas séries (Izzo et al. 1999) e que os jovens tenham

mais autonomia, Benner, Boyle e Sadler (2016) ressaltam que os pais continuam desempenhando um papel crucial nessa fase da vida, pois os filhos estão tomando decisões educacionais que influenciarão suas trajetórias subsequentes.

Esse artigo contribui com a literatura ao trazer para a discussão acadêmica a relevância dos fatores observáveis e não observáveis pelos quais a escolaridade dos pais pode afetar o desempenho acadêmico dos filhos, além de mensurá-los. Geralmente, as pesquisas que investigam essa relação adicionam isoladamente os níveis de escolaridade do pai e da mãe, não levando em consideração a composição educacional dos dois conjuntamente e os mecanismos pelos quais essa composição influencia os *outcomes* dos filhos. Sendo assim, trabalhos anteriores podem estar superestimando os impactos educacionais advindos da escolaridade do pai e da mãe. Adicionalmente, buscou-se diferenciar ao máximo os estudantes entre si, controlando por um extenso conjunto de características. Isso possibilitou analisar o desempenho entre estudantes que possuíam renda, tamanho da família e condições domiciliares semelhantes e ainda que estudavam na mesma escola, mas apenas com diferenças na escolaridade dos pais.

Em termos de efeito bruto, verificou-se que os filhos de pai e mãe com ensino superior registraram pontuação em média 80,99% maior do que os filhos de pai e mãe sem esse nível de escolaridade na prova de Redação. Eles também se saíram melhores nas demais provas: Linguagens e Códigos (16,06%), Matemática (27,40%) e Objetivas (20,95%). Ao controlar por renda, tamanho da família, infraestrutura domiciliar e escola (efeito líquido), filhos de pais com nível superior continuaram apresentando melhores resultados, mas as magnitudes dos coeficientes reduziram consideravelmente. Como exemplo, em Matemática caiu de 27,40% para 2,72% e em Objetivas reduziu de 20,95% para 2,58%. Além disso, as provas mostraram padrões semelhantes em termos de impacto das combinações educacionais do pai e da mãe, onde o impacto de quando apenas a mãe tem esse nível educacional é inferior à quando apenas o pai tem nível superior. Além disso, o efeito de apenas um dos pais possuir ensino superior é menor do que quando ambos são formados. Esse resultado é verificado tanto para o efeito bruto quanto após adicionar controles (efeito líquido).

Embora o efeito líquido pareça relativamente pequeno (variando de 1% a 10%), ele pode ser crucial para que o estudante consiga ser selecionado em cursos cuja concorrência seja acirrada, e a decisão entre os classificados e não classificados ocorra em uma pequena margem de diferenças na pontuação. Assim, os fatores não observáveis relacionados a pais mais educados pode ser um diferencial para aqueles que conseguem ser selecionados para um curso superior. Como exemplo, a pontuação média do exame Enem em 2019 foi em torno de 592,9 pontos e, nesse caso, o efeito líquido oscilaria entre 5,92 a 59,29 pontos.

Analisar o efeito bruto da educação dos pais sobre os filhos permite ter uma visão agregada, porém parcial, uma vez que há um conjunto de fatores correlacionados com a educação dos pais que afetam a performance estudantil e que precisam ser levados em consideração. Os resultados mostram que, quando controlamos pelos quatro grupos de efeitos fixos, a contribuição da escolaridade dos pais cai mais de 80% para todas as provas. Ainda assim, resta um efeito líquido que pode estar relacionado a uma gama de fatores não observáveis, dentre eles a transmissão de valores, expectativas, culturas, carga genética, dentre outros. O envolvimento dos pais no ambiente escolar e familiar (envolvimento parental) também tem sido amplamente discutido internacionalmente e apontado como fundamental no sucesso escolar.

Por fim, compreender os canais pelo quais a escolaridade dos pais influencia o desempenho dos filhos é importante para a formulação de políticas públicas e desenvolvimento de estratégias que busquem reduzir as disparidades educacionais e as desigualdades de oportunidades. Grande parte desses mecanismos possui persistência intergeracional e quanto

mais rápido se consegue identificá-los, maior o potencial de mitigar a transmissão de desigualdades socioeconômicas entre gerações.

REFERENCIAS

ABDUL-HAMID, H. (2007). Assessing Argentina's preparedness for the knowledge economy: Measuring student knowledge and skills in reading, mathematical and scientific literacy with evidence from PISA 2000. *Well-being and Social Policy*, 3(2), 41-66.

ALMOND, D., & CURRIE, J. (2010). Human Capital Development before Age Five. NBER Working Paper No. 15827. *National Bureau of Economic Research*.

BARBOSA, W. DE F.; SOUSA, E. P. DE. (2014) Análise do Desempenho Educacional dos Estudantes Cearenses no Exame Nacional do Ensino Médio. In: XV Semana de Iniciação Científica URCA, 2012, Crato, Ceará, p. 1-24, 2014.

BARROS, R., FOGUEL, M., & ULYSSEA, G. (2006). Desigualdade de Renda no Brasil: Uma Análise da Queda Recente. IPEA, Rio de Janeiro.

BECKER, G. (1964). Human Capital. A Theoretical and Empirical Analysis, with Special Reference to Education. NBER: Chicago.

BECKER, G. S., KOMINERS, S. D., MURPHY, K. M., & SPENKUCH, J. L. (2018). A theory of intergenerational mobility. *Journal of Political Economy*, 126(S1), S7-S25.

BEHRMAN, J. R., & ROSENZWEIG, M. R. (2002). Does increasing women's schooling raise the schooling of the next generation? *American Economic Review*, 92(1), 323-334.

BENNER, A. D., BOYLE, A. E., & SADLER, S. (2016). Parental involvement and adolescents' educational success: The roles of prior achievement and socioeconomic status. *Journal of youth and adolescence*, 45(6), 1053-1064

BJÖRKLUND, A., & RICHARDSON, K. (2001). The educational attainment of adopted children born abroad: Swedish evidence. *Unpublished manuscript, University of Stockholm*.

BJÖRKLUND, A., & SALVANES, K. G. (2011). Education and family background: Mechanisms and policies. In *Handbook of the Economics of Education* (Vol. 3, pp. 201-247). Elsevier.

BLACK, S. E., DEVEREUX, P. J., & SALVANES, K. G. (2005). The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, 120(2), 669-700.

BOONK, L., GIJSELAERS, H. J., RITZEN, H., & BRAND-GRUWEL, S. (2018). A review of the relationship between parental involvement indicators and academic achievement. *Educational Research Review*, 24, 10-30.

BREDTMANN, J., & SMITH, N. (2018). Inequalities in educational outcomes: How important is the family?. *Oxford Bulletin of Economics and Statistics*, 80(6), 1117-1144.

CASTRO, M., EXPÓSITO-CASAS, E., LÓPEZ-MARTÍN, E., LIZASOAIN, L., NAVARRO-ASENCIO, E. & CHEN, Q. (2009). Family Background, Ability and Student Achievement in

Rural China—Identifying the Effects of Unobservable Ability Using Famine-Generated Instruments.

COLEMAN, J. S. (1966). Equality of Educational Opportunity. US GPO: Washington DC

CONTINI, D., DI TOMMASO, M. L., & MENDOLIA, S. (2017). The gender gap in mathematics achievement: Evidence from Italian data. *Economics of Education Review*, 58, 32-42.

CORAK, M., PIRAINO, P., & FERREIRA, F. H. (2016). The Inheritance of Employers and Nonlinearities in Intergenerational Earnings Mobility. In *Inequality and Growth: Patterns and Policy* (pp. 1-34). Palgrave Macmillan, London.

CROOK, C. J. (1995). The role of mothers in the educational and status attainment of Australian men and women. *Australian and New Zealand Journal of Sociology*, 31(2), 45–73

CUNHA, FLAVIO AND JAMES J. HECKMAN. 2007. “The Technology of Skill Formation”. *American Economic Review*, 97(2), 31–47.

CURI, A. Z., & MENEZES FILHO, N. A. (2013). Mensalidade escolar, background familiar e os resultados do Exame Nacional do Ensino Médio (ENEM).

DAVIS-KEAN, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of family psychology*, 19(2), 294.

HOYOS, R. E. D., ESPINO, J. M., & GARCÍA, V. (2012). Determinantes del logro escolar en México. Primeros resultados utilizando la prueba ENLACE media superior. *El trimestre económico*, 79(316), 783-811.

ENGLAND, P., & SRIVASTAVA, A. (2013). Educational differences in US parents' time spent in child care: The role of culture and cross-spouse influence. *Social Science Research*, 42(4), 971-988.

ENTWISLE, D. R. (2018). Children, schools, and inequality. Routledge.

EPSTEIN, J. L. (1987). Toward a theory of family-school connections: Teacher practices and parent involvement. In K. Hurrelmann, F. Kaufmann, & F. Losel (Eds.), *Social intervention: Potential and constraints* (pp. 121-136). New York: Degruyter.

EPSTEIN, J. L., & LEE, S. (1995). National patterns of school and family connections in the middle grades. In B. A. Ryan, G. R. Adams, T. P. Gullotta, R. P. Weissberg, & R. L. Hampton (Eds.), *The family-school connection: Vol. 2. Theory research and practice* (pp. 108-154). Thousand Oaks: Sage.

FIGUEIRÊDO, E., NOGUEIRAY, L., & SANTANAZ, F. L. (2014). Igualdade de Oportunidades: Analisando o papel das circunstâncias no desempenho do ENEM. *Revista Brasileira de Economia*, 68(3), 373-392.

GHAZVINI, S. D., & KHAJEHPOUR, M. (2011). Gender differences in factors affecting academic performance of high school students. *Procedia-Social and Behavioral Sciences*, 15, 1040-1045.

GLEWWE, P. W., HANUSHEK, E. A., HUMPAGE, S. D., & RAVINA, R. (2011). *School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010* (No. 17554). National Bureau of Economic Research.

GLICK, P., & SAHN, D. E. (2000). Schooling of girls and boys in a West African country: the effects of parental education, income, and household structure. *Economics of education review*, 19(1), 63-87.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do Enem 2020**. Brasília: Inep, 2020. Disponível em: <<http://portal.inep.gov.br/web/guest/microdados>>. Acesso em: 01 set. 2020.

IZZO, Charles V. et al (1999). A longitudinal assessment of teacher perceptions of parent involvement in children's education and school performance. *American journal of community psychology*, v. 27, n. 6, p. 817-839.

JACOBS, N., & HARVEY, D. (2005). Do parents make a difference to children's academic achievement? Differences between parents of higher and lower achieving students. *Educational studies*, 31(4), 431-448

JEYNES, W. H. (2015). A meta-analysis: The relationship between father involvement and student academic achievement. *Urban Education*, 50(4), 387-423.

KALMIJN, M. (1994). MOTHER'S occupational status and children's schooling. *American Sociological Review*, 59(2), 257-275.

KORRUP, S. E., GANZEBOOM, H. B. G., & LIPPE, T. V. D. (2002). Do mothers matter? A comparison of models of the influence of mother's and father's education and occupational status on children's educational attainment. *Quality and Quantity*, 36(1), 17-42.

LAFORTUNE, J., & LEE, S. (2014). All for one? Family size and children's educational distribution under credit constraints. *American Economic Review*, 104(5), 365-69.

MARBUAH, D. A. (2016). Influence of Parental Income and Educational Attainment on Children's Years of Schooling: Case of Ghana.

MARKS, G. N. (2008). Are father's or mother's socioeconomic characteristics more important influences on student performance? Recent international evidence. *Social Indicators Research*, 85(2), 293-309.

MARSHALL, J. H., & SORTO, M. A. (2012). The effects of teacher mathematics knowledge and pedagogy on student achievement in rural Guatemala. *International Review of Education*, 58(2), 173-197.

MELO, L. DE, & ARAKAWA, V. H. (2016). Existe desigualdade regional na relação entre background familiar e desempenho escolar dos filhos? Evidências para as grandes regiões do Brasil. *Anais*.

MENDES, B. D., & KARRUZ, A. P. (2016). Background familiar, desigualdade regional e o desempenho no Exame Nacional do Ensino Médio (ENEM). *Belo Horizonte: Universidade Federal de Minas Gerais (UFMG)*, 1-25.

MENEZES-FILHO, N. A. (2007). *Os determinantes do desempenho escolar do Brasil* (pp. 1-31). IFB.

- MO, Y., & SINGH, K. (2008). Parents' relationships and involvement: Effects on students' school engagement and performance. *RMLE online*, 31(10), 1-11.
- NIEDERLE, M., & VESTERLUND, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, 24(2), 129-44.
- NIETO, S., & RAMOS, R. (2014). Decomposition of differences in PISA results in middle income countries. Working Paper 2014/08, p.38. Barcelona.
- NOBLE, K. G., HOUSTON, S. M., BRITO, N. H., BARTSCH, H., KAN, E., KUPERMAN, J. M., ... & SCHORK, N. J. (2015). Family income, parental education and brain structure in children and adolescents. *Nature neuroscience*, 18(5), 773.
- OREOPOULOS, P., & SALVANES, K. G. (2010). How Large are Returns to Education? Hint: Money isn't Everything. *Journal of Economic Perspectives*.
- PALERMO, G.A., SILVA, D.B.N. E NOVELLINO, M.S.F (2014). Fatores associados ao desempenho escolar: uma análise da proficiência em matemática dos alunos do 5º ano do ensino fundamental da rede municipal do Rio de Janeiro. *Revista Brasileira de Estudos de População*, 31(2), 367-394.
- PLUG, E. (2004). Estimating the effect of mother's schooling on children's schooling using a sample of adoptees. *American Economic Review*, 94(1), 358.
- POVEY, J., CAMPBELL, A. K., WILLIS, L. D., HAYNES, M., WESTERN, M., BENNETT, S. & PEDDE, C. (2016). Engaging parents in schools and building parent-school partnerships: The role of school and parent organisation leadership. *International Journal of Educational Research*, 79, 128-141.
- QIN, X., WANG, T., & ZHUANG, C. C. (2016). Intergenerational transfer of human capital and its impact on income mobility: Evidence from China. *China Economic Review*, 38, 306-321.
- REIS, M. C., & RAMOS, L. (2011). Escolaridade dos pais, desempenho no mercado de trabalho e desigualdade de rendimentos. *Revista Brasileira de Economia*, 65(2), 177-205.
- RIVKIN, S. G., HANUSHEK, E. A., & KAIN, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- RUMBERGER, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American educational Research journal*, 32(3), 583-625.
- SAMPAIO, B., & GUIMARÃES, J. (2009). Diferenças de eficiência entre ensino público e privado no Brasil. *Economia Aplicada*, 13(1), 45-68.
- SAMPAIO, B., SAMPAIO, Y., DE MELLO, E. P., & MELO, A. S. (2011). Desempenho no vestibular, background familiar e evasão: evidências da UFPE. *Economia Aplicada*, 15(2), 287-309.
- SANTOS, M. DOS M.; MARIANO, F. Z., & COSTA, E. M. (2018). Efeitos da Educação dos Pais sobre o Rendimento Escolar dos Filhos via Mediação das Condições Socioeconômicas. 46º Encontro Nacional de Economia (ANPEC 2018) Área 12-Economia Social e Demografia Econômica.

- SCHEEREN, L., DAS, M., & LIEFBROER, A. C (2017). Intergenerational transmission of educational attainment in adoptive families in the Netherlands. *Research in Social Stratification and Mobility*, 48.
- SCORZAFAVE, L. G., & FERREIRA, R. A. (2011). Desigualdade de proficiência no ensino fundamental público brasileiro: Uma análise de decomposição. *Revista Economia*, 12(2), 337-359.
- SCOTT-JONES, D. (1995). Parent-child interactions and school achievement. In B. A. Ryan, G. R. Adams, T. P. GULLOTTA, R. P. WEISSBERG, & R. L. HAMPTON (EDS.), *The family-school connection: Vol. 2. Theory, research, and practice* (pp. 75-109). Thousand Oaks: Sage.
- DE FARIAS SOUZA, Wallace Patrick Santos; DE OLIVEIRA, Victor Rodrigues; ANNEGUES, Ana Cláudia. Background familiar e desempenho escolar: uma abordagem não paramétrica. *Pesquisa e Planejamento Econômico*, v. 48, n. 2, 2018.
- TEACHMAN, J. D. (1987). Family background, educational resources and educational attainment, *American Sociological Review*, 52, 548–557.
- USEEM, E. L. (1992). Middle schools and math groups: Parents' involvement in children's placement. *Sociology of education*, 65, 263-279.
- VOYER, D., & VOYER, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological bulletin*, 140(4), 1174.
- YANG, J., & QIU, M. (2016). The impact of education on income inequality and intergenerational mobility. *China Economic Review*, 37, 110-125.
- WANG, Y., DENG, C., & YANG, X. (2016). Family economic status and parental involvement: Influences of parental expectation and perceived barriers. *School Psychology International*, 37(5).
- WANG, M. T., & SHEIKH-KHALIL, S. (2014). Does parental involvement matter for student achievement and mental health in high school? *Child development*, 85(2), 610-625.
- WILDER, S. (2014). Effects of parental involvement on academic achievement: A meta synthesis. *Educational Review*, 66(3), 377-397.
- WILLIAMS, P. (1980). Adolescent identification and academic achievement: reporting the awareness of similarity to role models, *Journal of Youth and Adolescence*, 9(4), 315–321.
- WITTEVEEN, D., & ATTEWELL, P. (2017). Family background and earnings inequality among college graduates. *Social Forces*, 95(4), 1539-1576.
- WOESSMANN, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *Journal of Economic Perspectives*, 30(3), 3-32.

ANEXO A

Quadro A1: Descrição das Áreas de Conhecimento e Componentes Curriculares do Enem 2017

Tipo	Áreas	Disciplinas	Pontuação e Número de Questões
Texto Argumentativo (1000 pontos)	Redação	-	C1 (200 pontos): Demonstrar domínio da modalidade escrita formal da língua portuguesa.
			C2 (200 pontos): Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa.
			C3 (200 pontos): Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista.
			C4 (200 pontos): Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação.
			C5 (200 pontos): Elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos.
Objetivas (180 questões)	Linguagens e Códigos e suas Tecnologias	Artes, Literatura, Língua Portuguesa, Educação Física, Língua Estrangeira (Inglês ou Espanhol) e Tecnologias da Comunicação e Informação.	45 questões
	Matemática	Álgebra e Geometria	45 questões
	Ciências da Natureza e suas Tecnologias	Química, Física e Biologia	45 questões
	Ciências Humanas e suas Tecnologias	História, Geografia, Filosofia e Sociologia	45 questões

Elaboração: própria dos autores com base no INEP.

ESCOLAS EM TEMPO INTEGRAL E DESEMPENHO NO ENEM: UMA AVALIAÇÃO DE IMPACTO PARA O ESTADO DO CEARÁ

Edward Costa Martins*

Gerrio dos Santos Barbosa**

Vitor Hugo Miro Couto Silva***

Helson Gomes de Souza****

Luciane dos Santos Jácome*****

RESUMO

O ensino integral é uma política que vem ganhando cada vez mais destaque no que se refere ao desempenho da educação no Estado do Ceará, principalmente quando se trata da sua aplicação no ensino médio. Baseado nisso, o objetivo deste trabalho é mensurar o efeito das Escolas em Tempo Integral (ETI) do Estado do Ceará sobre as notas do Enem referentes ao ano de 2017. Para tanto, foram utilizadas informações do Censo Escolar e da prova do Enem para a construção da base de dados utilizada. A partir de um conjunto de técnicas de quase experimento, os resultados demonstraram que as ETI possuem um efeito positivo sobre as notas das provas de todas as linhas de estudo abordadas no exame nacional do ensino médio. O principal destaque dos resultados está relacionado ao efeito da política sobre as notas de Redação, sendo que, a nota dos alunos das escolas integrais é, em média, cerca de 97 pontos superior à nota dos alunos que frequentaram escolas regulares em 2017. Assim, os resultados encontrados neste trabalho fornecem indícios de que o ensino em tempo integral melhora o desempenho dos alunos do ensino médio do estado do Ceará na prova do Enem.

Palavras-chave: Escolas em Tempo Integral; Prova do Enem; Ensino médio;

Classificação JEL: I20; I25; I26; I28.

Área 2 - Economia Social.

ABSTRACT

Full-time education is a policy that is gaining more and more prominence concerning the performance of education in the state of Ceará, especially when applied to high school. Based on this, the objective of this work is to measure the effect of the Full-Time Schools in the state of Ceará on the Enem test scores for the year 2017. For this purpose, we use information from the School Census and the Enem test to build the dataset. From a set of quasi-experiment techniques, the results showed that the Full-Time Schools have a positive effect on the test scores of all lines of study covered in the Enem test. The main highlight of the results is related to the effect of the policy on the Writing scores, so that, on average, the scores in Writing of the students from Full-Time Schools was 97 points higher than the scores of students from regular schools in 2017. Thus, the results found in this work provide evidence that full-time education increases the performance of high school students of the state of Ceará in the Enem exam.

Key Words: Full-Time Schools; Enem exam; High school.

JEL Classification: I20; I25; I26; I28.

* Doutor. Programa de Pós-Graduação em Economia Rural (PPGER-UFC). E-mail: edwardcosta@ufc.br.

** Doutorando em Economia Aplicada - UFPB. E-mail: gerriosantos@gmail.com. Tel.: (85) 989310682.

*** Doutor. Programa de Pós-Graduação em Economia Rural (PPGER-UFC). E-mail: vitormiro@gmail.com.

**** Doutorando em Economia Aplicada - UFPB. E-mail: helson.g.souza@gmail.com.

***** Doutoranda. Programa de Pós-Graduação em Economia Rural (PPGER-UFC).

1 INTRODUÇÃO

A educação é o principal fator que determina o crescimento econômico, o emprego e os ganhos futuros dos indivíduos (WOESSMANN, 2016). A compreensão desses efeitos é ponto chave na formulação de políticas públicas efetivas, sobretudo nesse novo contexto de diversas reformas no ensino brasileiro. De acordo com Filho e Pessoa (2010), a escola pública de ensino tradicional (regular) perdeu identidade — pois há uma nova realidade que demanda outras funções das instituições que não somente a aplicação da grade curricular formal — exigindo mais flexibilidade na formação e compromisso com a qualidade da formação, tendo em vista os benefícios sociais e econômicos aos quais os indivíduos necessitam no seu ciclo de vida.

No Brasil 1, há um esforço sendo tomado em todas as esferas da administração pública na tentativa de aumentar o número de escolas com ensino em tempo integral, buscando desenvolver não somente as capacidades cognitivas dos estudantes, mas principalmente, transformar o ambiente social no entorno dessas escolas, que na maioria das vezes são suscetíveis a altas vulnerabilidades sociais e econômicas. Em 2016, o Governo Federal (por meio do Fundo Nacional de Desenvolvimento - FNDE) lançou o Programa de fomento as escolas de ensino em tempo integral, liberando R\$ 406 milhões em recursos para apoiar os estados. No ano de 2018, houve uma ampliação para 967 escolas atendidas em todo o país, representando um aumento de 87% comparado ao ano de 2017. Além disso, as matrículas em escolas públicas de tempo integral no ensino médio aumentaram de 6,7% em 2016 para 8,4% em 2017.

No contexto de pioneirismo histórico, a concretização do ensino em tempo integral foi possível por intermédio do baiano Anísio Teixeira, que deu destaque a esse movimento em 1950, criando um complexo de quatro escolas e uma escola-parque em Salvador. Na década seguinte, Anísio replicou um modelo similar de ensino em Brasília, que tinha uma jornada escolar de 8 horas diárias, sendo as aulas complementares relacionadas a cultura, esporte, recreação e iniciação ao trabalho. No Estado de São Paulo, criou-se o Centros Vocacionais em 1962 para alunos de ambos os sexos e de 11 a 13 anos, com grade curricular organizada em 4 anos. Já no Estado do Rio de Janeiro, em 1985, Darcy Ribeiro criou as primeiras escolas brasileira de dia completo (Centros Integrados de Educação Pública - CIEPs), com 9 horas diárias para alunos da 1ª a 8ª série. Em 1986, o Estado de São Paulo criou o Programa de Formação Integral da Criança (Profic), com o objetivo de manter a criança na escola no momento em que os pais trabalhavam. Em relação a criação de Programas nacionais, no início da década de 1990 foram criados os Centros Integrados de Apoio à Criança (CIACs) e Centros de Atenção Integral à Criança (CAICs) pelos governos Collor de Melo e Itamar Franco, respectivamente (MILITÃO; KIILL, 2015).

No que diz respeito as legislações nacionais garantindo aplicação de ensino em tempo integral, pode-se resumir a LDB/96, PNE I, Fundeb, PME e PNE II.

No Estado do Ceará, o processo de criação de políticas direcionadas a aumentar o tempo escolar dos alunos de ensino médio ocorreu no ano de 2008, com as Escolas de Estaduais de Ensino Profissionalizante (EEEP), a quais combinam ensino em tempo integral e educação profissional. Em 2016, o Estado iniciou a Escola em Tempo Integral (ETI), que foi implementada por lei estadual em 2017, as quais buscam em suas diretrizes proporcionar aos filhos de trabalhadores uma formação que respeite seus potenciais, direitos de aprendizagem e desenvolvimento, além de desenvolver dimensões físicas, afetivas, cognitivas, intelectuais e éticas do ser humano, de tal forma que reduza as vulnerabilidades sociais e econômicas do Estado.

Para Bernardoni (2008) o processo de avaliação de um programa ou política constitui uma importante etapa do ciclo de políticas públicas, mas Figueiredo e Figueiredo (1986) afirmam que a constatação de que ocorreram mudanças não é por si só suficiente para concluir o sucesso de um programa, sendo então necessário demonstrar que as mudanças não ocorreriam (de forma total ou parcialmente) sem sua implementação, ou seja, é necessário fazer seu monitoramento e avaliação com intuito de utilizar de forma mais eficiente os recursos existentes. Assim, ao analisar um programa, tem que considerar não só seu trabalho imediato, mas entendê-lo como a possibilidade de melhorar a compreensão da realidade social e orientar a análise, formulação e implementação de tal política (JANNUZZI, 2002), podendo levar a um novo direcionamento.

Nesse sentido, nas últimas décadas os investimentos em educação no Brasil aumentaram consideravelmente e, o governo federal passou a acompanhar por meio de avaliações de proficiência tanto a educação básica, com base no Índice de Desenvolvimento da Educação Básica (IDEB) e da Prova Brasil, quanto a educação superior por intermédio do Exame Nacional de Desempenho de Estudantes (ENADE), para verificar o retorno dos investimentos nas políticas educacionais e analisar se os resultados obtidos são positivos. A necessidade do investimento em educação se mostra evidente, não apenas quanto ao indivíduo para a redução das desigualdades de oportunidade, mas também em nível nacional, para a determinação do produto e do crescimento, tendo os projetos de escola de tempo integral como uma política de melhoria da qualidade educacional (SIMÕES; CIRINO; CASSUCE, 2014).

Os Programas direcionados ao ensino de tempo integral ainda são escassos no país, bem como a literatura que busque mostrar sua eficácia em promover melhorias nos mais diversos indicadores educacionais, além do mais, os principais responsáveis pelo ensino médio são as unidades federativas. Diante disso, a presente pesquisa tem como motivação a avaliação de impacto das ETI's sobre as notas do Enem de 2017 para o Estado do Ceará, com o objetivo de mensurar a efetividade da política fazendo uma breve discussão relacionado aos resultados de outras políticas semelhantes. Para atingir os objetivos serão usadas técnicas econométricas de inferência causal, realizando comparações de eficiência dos métodos e dando consistência aos resultados encontrados.

A presente pesquisa apresenta um diferencial e contribuição na análise de políticas públicas educacionais, tendo em vista que não foram encontrados pesquisas na literatura recente de avaliações de Programas de escolas em tempo integral utilizando uma abordagem com método de entropia, a qual se destaca em relação aos métodos tradicionais. Outra contribuição importante é a formação de literatura para os formadores de políticas públicas, sobretudo de políticas educacionais do ensino médio.

Este estudo está organizado em mais quatro seções, além desta introdução. A segunda seção descreve a revisão de literatura, seguido da metodologia aplicada. Na quarta seção se encontra os resultados e análises e finalizando com as considerações finais.

2 REVISÃO DE LITERATURA

Nesta seção, decorre-se a respeito da educação integral do Brasil, assim como a implantação da mesma no Estado do Ceará. Além disso, uma breve explanação sobre o Exame Nacional do Ensino Médio.

2.1 Escola de Tempo Integral

A idealização da educação integral surgiu em 1932 quando Manifesto dos Pioneiros da Educação Nova foi divulgado, com uma renovação no ensino, denominada Escola Nova que

defendia também a universalização, a laicidade e a gratuidade da escola pública (DUTRA, 2014), mas foi no fim da década de 1950, na cidade de Salvador, na Bahia, quando Anísio Teixeira criou a Escola Parque, sendo então, a primeira prática de educação integral, voltada para crianças nas séries iniciais da educação básica, incluindo no processo de escolarização diversas ações de socialização e práticas de aprendizagem não formais, como atividades culturais e de lazer.

Existe, pelo menos, quatro conceitos acerca da escola de tempo integral, como destaca (CAVALIERE, 2002), sendo a primeira, a visão assistencialista, que coloca a escola de tempo integral como uma organização que visa suprir as deficiências de aprendizagem de pessoas desprivilegiadas. Já na visão autoritária, a escola de tempo integral é vista como uma instituição que tira os estudantes das ruas, prevenindo-os da violência. A visão democrática considera a escola com jornada prolongada como uma organização que prepara os alunos, de forma crítica, para o mundo. Para finalizar, a recente concepção, é a visão multissetorial, que independe do tempo de escola. Como Gonçalves (2006) afirmou, a visão multissetorial é a mais apresentada pelos educadores e a educação integral vai além da concepção cognitiva, ou seja, o sujeito está, também, inserido em um contexto de relações sociais e afetivas. Além do mais, deve-se pensar nos estudantes como sujeitos inteiros, considerando seus interesses, assim como, o diálogo entre educadores e educandos é fundamental, pois a ampliação da jornada escolar só faz sentido se for de encontro com a aplicação das oportunidades que promovem a melhoria do aprendizado.

Segundo Cavaliere (2002), houve a necessidade de uma nova identidade para a escola pública, a influência de políticas públicas, como o Bolsa Família, que exige a permanência dos jovens na escola, a retirada da classe média urbana da rede pública para a rede privada e a privatização do ensino. Com isso, Cavaliere (2002) ainda explica que a ampliação da jornada escolar diária pode ser entendida e justificada como uma forma de melhorar a ação escolar sobre os resultados dos discentes, de adequar as novas configurações familiares da vida urbana, ou ainda, em decorrência da mudança na concepção de educação escolar, ou seja, do papel da escola na vida e na formação dos estudantes. No âmbito federal, existem dois programas de ensino em tempo integral: o Mais Educação a nível fundamental e o Ensino Médio Inovador (ProEMI) para o nível médio, sendo que ambos têm o propósito de deixar o ensino brasileiro mais dinâmico por meio de ações como o acompanhamento pedagógico às áreas do conhecimento, assim como estimular o desenvolvimento de atividades artísticas, culturais, esportivas, de lazer de comunicação, de cultura digital, entre outras que possam desenvolver as múltiplas dimensões humanas (cognitivas, psicomotoras, socioafetivas, estética, ética).

No Estado do Ceará, a política de educação de escola de tempo integral foi iniciada a partir do Programa Brasil Profissionalizado, em 2008, sendo instituídas as Escolas Estaduais de Educação Profissional (EEEP). As EEEP's têm em seu currículo cursos técnicos e profissionalizantes em diversas áreas, que buscam uma formação pessoal e convívio em sociedade, onde além de concluir o ensino médio básico, o estudante se profissionaliza em áreas que o voltem para o mercado de trabalho (MARIANO, 2016). O modelo integrado possibilita a centenas de alunos a qualificação para ingressar no mercado de trabalho ao mesmo tempo que são habilitados a concorrer a uma vaga na universidade.

A partir do 2016, o Governo Estadual do Ceará vem implantando o Programa Escola em Tempo Integral (ETI), sendo uma estratégia que defende que os estabelecimentos públicos consigam proporcionar aos filhos de trabalhadores uma formação integral, respeitando seus potenciais, direitos de aprendizagem e desenvolvimento, ou seja, as escolas de tempo integral buscam ampliar o processo educativo visando novas possibilidades, permitindo aos sujeitos envolvidos a vivência de experiências metodológicas diferenciadas.

A escola em tempo integral do Ceará está alicerçada à meta 6 do Plano Nacional de Educação (PNE), que visa oferecer educação em tempo integral, no mínimo, 50% das escolas

públicas, atendendo pelo menos 25% dos alunos da educação básica, e à meta 6 do Plano Estadual de Educação (PEE), que busca, até 2024, oferecer, no mínimo 50% das escolas públicas e instituições de educação infantil, atendendo, pelo menos, 25% dos estudantes da educação básica. O Programa de ETI teve seu marco legal de ensino a partir da Lei Nº 16.287, de 20 de julho de 2017, apesar de já funcionar desde 2016.

O projeto pedagógico das escolas de tempo integral cearenses é estruturada por três dimensões fundamentais: a escola deve ser concebida como comunidade de aprendizagem, com a participação e a interação da sociedade; a aprendizagem cooperativa deve ser o método pedagógico estruturante, em que a aprendizagem acontece a partir da efetiva interação com seus colegas; o protagonismo estudantil é um princípio imperativo para a proposta de ensino médio, onde o jovem estimulado não é apenas o receptor de conhecimento, mas também, aquele que é capaz de fazer escolhas e ajudar seus pares a superarem os desafios de forma criativa. Cada uma dessas dimensões deve ser trabalhada de forma intrínseca, compreendendo-as como inseparáveis.

2.2 Exame Nacional do Ensino Médio (Enem)

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) criou, em 1998,¹ o Exame Nacional do Ensino Médio² (Enem), cujo principal objetivo é possibilitar uma alusão para uma autoavaliação, por meio de competências e habilidades que o compõem, avaliando o desempenho do estudante ao fim da escolaridade básica. Servindo, também, como uma forma alternativa ou complementar às seleções de ingresso ao nível superior. É um exame individual e de caráter voluntário, realizado anualmente aos concluintes e egressos de ensino médio, sendo um importante método de avaliação, por fornecer informações atuais e realistas da educação brasileira.

Após sua criação, o Enem passou por algumas reformulações, onde em 2004 sua nota passou a ser usada como critério na concessão de bolsas do Programa Universidade para Todos (Prouni), que oferece bolsas de estudo integral e parciais em instituições de ensino superior privadas. Contudo, a maior reforma aconteceu em 2009, quando foi criado o Sistema de Seleção Unificado (SiSU), onde instituições públicas de ensino superior oferecem vagas para candidatos participantes do exame, ou seja, o Enem passou a substituir provas de vestibulares. Ademais, o exame passou a cumprir para a certificação do Ensino Médio para aqueles que não terminaram a escolarização básica, pessoas privadas de liberdade e jovens sob medida socioeducativa. Direito estabelecido pela Lei das Diretrizes e Bases da Educação Nacional – Lei nº 9.394 de 20 de dezembro de 1996, artigo 38.

Desde sua criação até 2008, a prova era constituída por 63 questões de múltipla escolha, onde todas as perguntas tinham o mesmo valor, portanto para se obter a nota global multiplicava-se o número de questões corretas por 100 e em seguida, dividia por 63. Atualmente, o Enem possui 180 questões distribuídas em quatro áreas de conhecimento: Ciências Humanas e suas Tecnologias (CH); Ciências da Natureza e suas Tecnologias (CN); Linguagens, Códigos e suas Tecnologias (LC); Matemática e suas Tecnologias (MaTec) e uma redação. Por meio de representações do Ministério da Educação (MEC) e da Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior (ANDIFES), o conteúdo do exame está ordenado em cinco vertentes cognitivas: 1. dominar linguagens; 2. compreender fenômenos; 3. enfrentar situações-problema; 4. construir argumentação; 5. Elaborar propostas.

¹ O texto completo em: <http://portal.mec.gov.br/component/tags/tag/41541>.

² Esse resumo sobre o Enem foi realizado a partir do site. <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>

Portanto, exames como Enem contém informações necessárias para se avaliar como políticas públicas educacionais têm impactado o desempenho acadêmico dos estudantes, bem como se as mesmas estão cumprindo seus objetivos ou se são necessária novas reformulações.

2.3 Efeitos do Programa Escola de Tempo Integral

Na literatura internacional destacam-se Lee e Barro (2001), Cerdan-Infantes e Vermeersch (2007) e Bellei (2009). O objetivo da pesquisa de Lee e Barro (2001) foram crianças americanas com menor idade, onde se buscava verificar se as crianças que estudavam no ensino primário em tempo integral aprendiam mais do que aquelas que frequentavam escolas de tempo regular e, o resultado se mostrou positivo em crianças que permaneciam na escola por mais tempo, obtendo ainda maior capacidade para ler e escrever, além de um melhor desenvolvimento em matemática.

Por outro lado, Cerdan-Infantes e Vermeersch (2007) e Bellei (2009) avaliaram o impacto do programa Escola de Tempo Integral sobre crianças uruguaias da sexta série, usando como metodologia o propensity score matching. Os resultados indicaram que o desempenho dos alunos aumentaram em 0,07 e 0,04 em matemática e idioma, respectivamente, a cada ano que permanecem no programa. Bellei (2009) a partir do método de diferença em diferenças, comparou estudantes do 10º, nos anos de 2001 e 2003 e mostrou que a Escola de Tempo Integral Chilena possui resultados favoráveis do programa em relação ao desempenho dos discentes, sendo 3,5 pontos a mais para matemática e 2,5 pontos para leitura.

Em relação a literatura nacional, destaca-se a pesquisa de Aquino (2011), que avaliou o programa Escola de Tempo Integral da rede pública do Estado de São Paulo, tendo como objetivo verificar se a ampliação da jornada escolar teria impacto sobre o rendimento dos alunos da 8ª série nas disciplinas de português e matemática. Usando dados do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP) de 2007 e 2008 e do Censo Escolar de 2005, 2007 e 2009, o estudo mostrou que não houve efeito para matemática, porém mostrou-se significativo para português. Apesar de não haver ganhos na proficiência, a autora comenta sobre a importância da escola integral por diminuir a ocorrência do trabalho infantil e aumento da interação social.

Na mesma perspectiva, Soares et al. (2014) utilizaram dados do Programa de Avaliação da Educação Básica de Minas Gerais (PROERB) de 2006 a 2009, avaliaram a Escola de Tempo Integral com técnicas de score de propensão e modelos de multiníveis hierárquicos, comparando escolas que aderiram ao programa e aquelas que não aderiram. Nos modelos hierárquicos, os resultados se mostraram significativos, principalmente para aqueles que apresentaram menor rendimento. E nos modelos de escores de propensão, as escolas participantes, mesmo com notas menores, diminuíram a diferença que havia em um período anterior, o que condiz com o objetivo do programa, pois sua implantação inicial tem por finalidade atender as escolas com alunos mais vulneráveis e baixa proficiência.

Buscando avaliar a educação Profissional e Tecnológica (EPT) de nível médio sobre o desempenho e a inserção no mercado de trabalho, formando um grupo de controle com as escolas regulares, Araújo, Chein e Pinto (2018) utilizaram dados do Enem 2009 e abordagem PSM. Os autores estimaram os efeitos sobre as notas do Enem de alunos que cursaram EPT comparado aos alunos de ensino regular (em tempo parcial), sendo os resultados estatisticamente significantes e divididos para as várias dependências administrativas. Seu principal achado apontou que alunos das EPT's obtiveram em Linguagens e Códigos e Redação entre 4,5 e 13,44 pontos, em média, acima dos alunos do grupo de comparação, respectivamente. Suas conclusões apontam para a possibilidade de seus resultados não serem causais devido a fatores de autoseleção e não habilidades não observáveis. Outros resultados

podem ser relacionados ao mercado de trabalho, mostrando que há mais probabilidade de um aluno de escola EPT estar empregado naquele ano.

Portanto, diante da literatura apresentada estudos de impactos das escolas de tempo integral sobre o desempenho acadêmico ainda são escassos. Assim, este estudo trará para a literatura uma análise por meio de técnicas de avaliação de impactos apresentadas na próxima seção.

3 PROCEDIMENTOS METODOLÓGICOS

3.1 Dados e Variáveis

Os dados utilizados são do Censo Escolar (2017 e 2018) e Enem (2017). Ambas as bases podem ser encontradas no Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). O ano de 2017 é escolhido devido ter um maior número de escolas na modalidade de ETI comparado a 2016, que funcionou como um ano teste. No Censo Escolar de 2017 são extraídas informações das escolas e tempo de duração do ensino (usada na estratégia de identificação), enquanto a base do Enem é usada fornece as variáveis para construir o modelo (usada na estratégia empírica).

No Quadro 1 são mostrados os nomes atribuídos a cada variável utilizada, bem como as suas respectivas descrições detalhadas e as fontes de extração. De certa forma, praticamente todas essas variáveis já são usadas recursivamente em pesquisas relacionadas a economia da educação — no entanto — a escolha teve como base os estudos recentes de Araújo, Chein e Pinto (2018) e Mariano (2016). Contudo, a pesquisa contribui na criação de três variáveis ainda não exploradas na literatura, sendo o efeito deslocamento, o efeito cansaço e o efeito migração. A primeira, busca investigar o efeito do indivíduo estudar em uma escola de um município que não seja o de residência, os quais podem ser positivos na participação do Programa (pois o indivíduo que se desloca de uma cidade para outra em busca de ensino integral tem um elevado nível de autoseleção) e pode ser negativo nos resultados (pelo fato de que, comparado aos seus pares que residem e estudam na mesma cidade, você terá menos tempo e mais fadiga). A segunda tem o objetivo de averiguar se o município de residência do estudante é diferente do município de nascimento, podendo ter um efeito positivo na participação do Programa (essa variável se diferencia do efeito deslocamento, devido o indivíduo que migrou permanecer residente, uma limitação é que o indivíduo pode ter migrado há muito tempo), acredita-se que esse efeito fosse ainda maior se o indivíduo tivesse migrado no ano de 2017 e o motivo fosse a ETI. A terceira analisa o possível efeito do indivíduo sair de sua cidade de residência para fazer prova do Enem em outra cidade (acumulando fadiga e perda de tempo no dia da prova), sendo tais efeitos observados somente na variável de resultado (nas notas das provas do Enem) e não influencia a participação do aluno no ETI.

Quadro 1: Descrição e fonte das variáveis

Variável	Descrição	Fonte
Variáveis Dependentes		
Q X bQ R WD bFQ	Valor referente à nota de Ciências da Natureza da prova do ENEM	(INEP, 2020)
Q X bQ R WD bFK	Valor referente à nota de Ciências Humanas da prova do ENEM	(INEP, 2020)
Q X bQ R WD bOF	Valor referente à nota de Linguagens e Códigos da prova do ENEM	(INEP, 2020)
Q X bQ R WD bP W	Valor referente à nota de Matemática da prova do ENEM	(INEP, 2020)
Q X bQ R WD bUHG	Valor referente à nota de matemática da prova do ENEM	(INEP, 2020)
Variáveis que Afetam a Participação no ETI		
hnhlwG hvarfcp hqwr	1 se o indivíduo estuda em município diferente do que ele reside e zero caso contrário	(INEP, 2020)
hnhlwP ljucfdr	1 se o indivíduo estuda na mesma unidade da Federação que nasceu e zero caso contrário	(INEP, 2020)
vxshuSdl	1 se o pai do indivíduo possui nível superior completo e zero caso contrário	(INEP, 2020)

vshuP dh	1 se a mãe do indivíduo possui nível superior completo e zero caso contrário	(INEP, 2020)
p nudgruhMF dvj	Número de moradores do domicílio	(INEP, 2020)
vdcluhP lqj r	Renda mensal domiciliar em salários mínimos	(INEP, 2020)
hmfSxeadfdHP	1 se existe se o indivíduo cursou todo o ensino médio em escola pública e zero caso contrário	(INEP, 2020)
Variáveis que Afetam os Resultados		
p dvfxcdqr	1 se o indivíduo é do sexo masculino e zero caso contrário	(INEP, 2020)
hwdf lyk	1 se o indivíduo é solteiro e zero caso contrário	(INEP, 2020)
gdrg hfcdlgr	1 se o indivíduo não declarou sua cor/raça e zero caso contrário	(INEP, 2020)
eudqfr	1 se o indivíduo é autodeclarado de cor branca e zero caso contrário	(INEP, 2020)
dp dhr	1 se o indivíduo é autodeclarado de cor amarela e zero caso contrário	(INEP, 2020)
sdgr	1 se o indivíduo é autodeclarado de cor parda e zero caso contrário	(INEP, 2020)
shwr	1 se o indivíduo é autodeclarado de cor preta e zero caso contrário	(INEP, 2020)
lqgljrd	1 se o indivíduo é autodeclarado indígena e zero caso contrário	(INEP, 2020)
hnhwF dqvdf	1 se o indivíduo realiza a prova do ENEM em município diferente do que ele estuda e zero caso contrário	(INEP, 2020)
hwfrcX uedqd	1 se a escola em que o indivíduo estuda está localizada na zona urbana e zero caso contrário	(INEP, 2020)
wjp WY	1 se existe TV na residência do indivíduo e zero caso contrário	(INEP, 2020)
wySnD wlg	1 se existe TV por assinatura na residência do indivíduo e zero caso contrário	(INEP, 2020)
wjp Fhxat	1 se se o indivíduo possui aparelho celular e zero caso contrário	(INEP, 2020)
wjp wrclr	1 se existe telefone fixo na residência do indivíduo e zero caso contrário	(INEP, 2020)
wjp F rp sxwgru	1 se existe computador na residência do indivíduo e zero caso contrário	(INEP, 2020)
wjp lqwhqhv	1 se existe internet na residência do indivíduo e zero caso contrário	(INEP, 2020)

Fonte: Elaboração própria dos autores.

Nota 1: As variáveis que afetam a participação também afetam o resultado do Programa.

3.2 Estratégia de Identificação

Com intuito de atingir os objetivos da presente pesquisa, buscou-se organizar uma base de dados das ETI do estado do Ceará para o ano de 2017 utilizando as bases públicas do Inep (Censo Escolar e Enem). Entretanto não há uma variável de ligação a nível de indivíduo nas duas bases, o que impossibilita ligar as variáveis do Censo Escolar relacionadas ao aluno (como por exemplo, data de nascimento) com a base do Enem. Contudo, a base do Censo Escolar fornece a possibilidade de examinar quais são as escolas de tempo integral segundo a MEC³ (ou seja, que possuem acima de 420 minutos de aula no dia letivo), as que são profissionalizantes (que também possuem acima de 420 minutos de aula por dia letivo) e as que são de tempo parcial (que possuem menos de 420 minutos de aula por dia letivo). Por outro lado, há a possibilidade de ligar as bases por meio do código de identificação das escolas, sendo assim, as bases do censo são compiladas por escolas integrais e não integrais no ano de 2017 e, posteriormente, combinada a base dados do Enem a nível de indivíduo (essa base possui o código de identificação da escola cursada pelo indivíduo no ensino médio).

O grupo de tratamento é construído com as escolas que passaram a ter mais de 9 horas aula no ano de 2017. Já o grupo de controle é formado pelas escolas que em 2017 eram de tempo parcial, mas em 2018 (Censo Escolar) aparecem como ETI. Usa-se somente alunos concludentes do 3º ano do ensino médio em 2017, que frequentavam escolas públicas estaduais em atividade com ensino regular e com tempo de aula por dia letivo acima de 540 minutos (ou 9 horas aula dia letivo segundo a Lei Nº 16.287, de 20 de julho de 2017). O tempo é escolhido para tornar mais homogênea os grupos de tratados e não tratados, além do mais, as escolas que oferecem um tempo menor não apresentam uma quantidade significativa de alunos.

³ Definição encontrada no Decreto nº 7.083, de 27 de Janeiro de 2010, na Resolução nº 7 de 14 de Dezembro de 2010 e, mais recente, o MEC divulgou as novas diretrizes do ensino de tempo integral na Portaria nº 2.116 de 06 de Dezembro de 2019.

Os critérios para selecionar escolas de tempo integral foram estabelecidos legalmente em agosto de 2017 (Lei Nº 16.287, de 20 de julho de 2017), portanto, a lógica supõe que os critérios utilizados no início de 2017 para seleção das escolas de ETI foram os mesmos utilizados em 2016, ou seja, as escolas com 50% ou mais de alunos no Programa de Transferência de Renda do Governo Federal (Bolsa Família). Contudo, os critérios usados em 2016 são fortemente correlacionados com os critérios estabelecidos em 2017, não havendo assim, perda de generalidade.

3.3 Estratégia Empírica

3.3.1 Pareamento por Escore de Propensão

Para mensurar os efeitos das ETI sobre as notas do Enem é necessário que a amostra contenha informações dos indivíduos em duas ocasiões, sendo a primeira, a ocasião na qual o indivíduo foi tratado, ou seja, foi aluno de uma escola integral, e a segunda, denotando informações contrafactuais as quais demonstram como seria as características desse mesmo indivíduo, no mesmo período, caso ele não fosse tratado.

Para fundamentar matematicamente esses conceitos e como eles são usados para medir o efeito de um determinado fenômeno ou intervenção, este trabalho faz uso da notação utilizada por Hirano, Imbens e Ridder (2003) e que é recorrente na literatura que trata da avaliação de impacto. Inicialmente, considere uma amostra de tamanho Q . Para cada unidade l na amostra, $l = 1, 2, \dots, Q$, deixe W_l denotar a condição de tratamento do indivíduo l de maneira que, $W_l = 1$ indica que o indivíduo l foi tratado, enquanto os não tratados recebem valor zero para W_l . Considere $Y_l(0)$ como sendo a nota do Enem referente ao indivíduo l pertencente ao grupo de controle e $Y_l(1)$ como sendo a nota do Enem referente ao indivíduo l pertencente ao grupo de tratamento. Visto isso, observa-se o valor de $Y_l = W_l Y_l(1) + (1 - W_l) Y_l(0)$.

O que se busca com esse procedimento é encontrar o Efeito Médio do Tratamento sobre os Tratados - EMTT (em inglês, *Debiased Heteroskedasticity Weighted OLS*), o qual pode ser obtido por meio da seguinte

especificação:

$$EPT \equiv E[Y_l(1) - Y_l(0) | W_l = 1] \quad (1)$$

Para que o valor do EMTT seja interpretado sob um ponto de vista estatisticamente válido é preciso comparar os indivíduos que recebem o tratamento com este mesmo indivíduo em uma situação de não- tratamento. Dada a impossibilidade de observar o mesmo aluno tendo ele estudado em uma escola integral e em uma escola não integral, ao mesmo tempo, nós optamos por utilizar uma técnica de quase-experimento conhecida como pareamento. Uma versão bastante comum desta técnica é o pareamento por escore de propensão. Este mecanismo é fundamentado na hipótese de que é possível comparar um indivíduo tratado com um indivíduo não tratado desde que a probabilidade de tratamento desses dois indivíduos seja semelhante. Com isso, é possível que cada membro do grupo de tratamento possua um contrafactual no grupo de controle, que represente de maneira mais adequada possível o resultado que o indivíduo tratado teria obtido caso não fosse tratado.

Cabe destacar que algumas hipóteses são assumidas durante a construção do pareamento. A primeira delas é conhecida como ignorabilidade (ou independência condicional) e trata do fato de que as variáveis observáveis (X_j) possuem todas as informações sobre $Y_i(0)$ e $Y_i(1)$, o que proporciona uma situação de independência entre $Y_i(0)$ e $Y_i(1)$. Outra hipótese assumida nesse processo é a denominada hipótese de sobreposição, a qual indica que

a região formada pela área de dispersão das variáveis observáveis referentes às características dos indivíduos tratados representa as características dos indivíduos que estão no grupo de não tratado. Essa hipótese deve ser configurada de maneira a garantir que para cada indivíduo no grupo de controle, deve existir um par no grupo de tratamento, cujo resultado potencial seja um bom predictor do resultado desse indivíduo, caso ele fosse tratado, garantindo que as características dos tratados sejam representadas no grupo de controle e vice-versa.

Uma das maneiras mais comuns de se efetuar o pareamento e a consequente amostra de contrafactuais é por meio da utilização do pareamento por score de propensão, onde a probabilidade de cada indivíduo ser tratado é atribuída a um vetor de covariadas observáveis. Com isso, o pareamento dos grupos é feito por meio da semelhança entre essa probabilidade estimada ⁴. Formalmente, o escore de propensão é dado por:

$$S(X_i) = S[W_1 = 1 | X_i] \quad (2)$$

De maneira que $S(X_i)$ pode ser obtida por meio da utilização de um modelo para variável dependente binária. No caso do presente ensaio, a probabilidade de tratamento será calculada por meio da utilização de um modelo logit . De forma similar, será também aplicado o método do PDK, que ao contrário do PSM, usa uma função de distância não paramétrica para resolver o problema da dimensionalidade, sendo mais explicado em Cochran e Rubin (1973).

3.3.2 Balanceamento por Entropia

Considere que o efeito médio do tratamento sobre os tratados seja fielmente representado pela Equação 1, nesse caso, uma média contrafactual para a variável de interesse pode ser obtida de acordo com:

$$E[Y(0) | T = 1] = \frac{\sum_{i \in \{T=0\}} Y_i w_i}{\sum_{i \in \{T=0\}} w_i} \quad (3)$$

Em que w_i é o peso escolhido para cada unidade de controle. Os pesos são escolhidos de acordo com a seguinte regra de reponderação:

$$\min_{w_i} H(w) = h\left(\sum_{i \in \{T=0\}} w_i\right) \quad (4)$$

Sujeito às seguintes restrições de normalização do balanceamento:

$$\sum_{i \in \{T=0\}} w_i c_{ri}(X_i) = m_r \quad \text{com } r \in 1, \dots, R \quad \text{e} \quad \sum_{i \in \{T=0\}} w_i = 1 \quad \text{e} \quad (5)$$

$w_i \geq 0$ Para todo i dado que $T = 0$

Em que $k(\cdot)$ é a distância métrica e $\mathbb{E}_U(\cdot)$ descreve o conjunto de U restrições de balanceamento impostas sobre os momentos das covariadas do grupo de controle reponderado. As restrições de balanceamento garantem que os grupos ponderados correspondem de forma

⁴ Para mais detalhes, ver Rosenbaum e Rubin (1983)

precisa aos momentos especificados. Os pesos que resultam deste balanceamento podem ser passados para diversos padrões de modelos a serem estimados pelo pesquisador⁵.

O balanceamento, sendo feito por meio do escore de propensão ou da entropia, tem a finalidade de selecionar um grupo semelhante ao verdadeiro contrafactual dos indivíduos tratados. Tendo realizado esse procedimento, o HP WW pode ser obtido por meio do cálculo da Equação 1 com o banco de dados formado pelos indivíduos tratados e os seus respectivos contrafactuais selecionados no pareamento. Como o HP WW é uma média, este valor pode ser obtido por meio do coeficiente de inclinação da reta de regressão resultante da seguinte estimação:

$$|_1 = \alpha + \beta W_1 + \vartheta \Gamma_1 + \varepsilon \quad (6)$$

Em que $|_1$ é o resultado da nota do Enem, β representa o HP WW, Γ_1 é uma matriz de covariadas que extraem o valor de β que não está relacionado com o tratamento, ϑ mede o efeito de cada covariada sobre a nota do ENEM e ε é o termo de erro da estimativa. Usando os pesos calculados nas técnicas de pareamento, este trabalho estima a Equação 6 por meio de um estimador de mínimos quadrados ponderados.

4 RESULTADOS

Esta seção apresenta as estratégias de quase experimentos adotadas para atingir os objetivos da pesquisa, que é avaliar o impacto dos estudantes de ensino em tempo integral nas notas do Enem para o estado do Ceará no ano de 2017. Na Subseção 4.1 serão mostradas o pareamento por escore de propensão (PSM) e o balanceamento por entropia, enquanto a Subseção 4.2 confronta as estimações e discute seus resultados principais.

4.1 Pareamento por PSM e Reponderação por Entropia

O método de quase experimento mais utilizado na análise de impactos de políticas públicas em determinado público alvo é o PSM. As descrições na Tabela 1 mostram o comportamento dos grupos de tratado e não tratado após o pré-processamento dos dados. Nesse primeiro momento são utilizadas variáveis que afetam a participação do aluno no Programa ETI na busca em mitigar o viés de autosseleção. Antes do pareamento, as médias das variáveis relacionadas aos pais dos alunos terem ensino superior, apresentam-se ajustadas (ou seja, estatisticamente iguais a zero). Por outro lado, após ser realizado a correspondência pelo escore de propensão entre os grupos tratados e não tratados, todas as variáveis que afetam a autosseleção possuem diferenças de médias estatisticamente iguais a zero, indicando o bom ajuste das covariadas pelo uso do método PSM.

Tabela 1: Diferença de médias antes e após o pareamento por PSM.

Variável	Antes do Pareamento			Após do Pareamento		
	Não Tratado	Tratado	P-Valor Variável	Não Tratado	Tratado	P-Valor
efeitoDeslocamento	0,012	0,061	0,001	0,046	0,057	0,554
efeitoMigracao	0,198	0,133	0,005	0,122	0,134	0,695
superPai	0,270	0,285	0,618	0,260	0,286	0,493
superMae	0,281	0,304	0,455	0,305	0,305	1,000
moradoresCasa	4,299	4,468	0,076	4,538	4,462	0,584
salarioMinimo	0,705	0,38	0,000	0,374	0,382	0,857
escPublicaEM	0,969	0,985	0,071	0,989	0,985	0,704

Fonte: Elaboração própria dos autores.

⁵ Para maiores detalhes sobre o balanceamento com entropia, ler Hainmueller (2012).

De acordo com King e Nielsen (2019), o uso do PSM (e seus muitos refinamentos) não possibilita analisar efeitos causais e, que na verdade, esse método aumenta o desequilíbrio, a ineficiência, a dependência do modelo, o critério de pesquisa e o viés estatístico. Ademais, o estudo indica que quanto mais balanceado estão os dados devido aos descartes realizados pelo pareamento, será mais provável a degradação da inferência causal — o que foi chamado de Paradoxo do PSM. Por fim, os autores sugeriram⁶ que o MDM é um método de pareamento que não sofre desses problemas⁷. Entretanto, Um método recente foi criado por Hainmueller (2012), o qual busca balancear os dados por reponderação no pré-processamento. As principais vantagens⁸ desse método comparado ao PSM e MDM é seu balanceamento exato envolvendo os três momentos (ou até momentos mais altos), a retenção de informações (não há descarte — a reponderação garante a manutenção de toda a amostra), o não uso de formas iterativas⁹ (testando diversas formas funcionais para encontrar a que melhor se ajusta aos dados) para encontrar os pesos, a versatilidade em utilizar diversas modelagens padrões de estimações subsequente a reponderação (como regressões, diferenças de médias, etc.) e apresenta-se computacionalmente mais ágil (dado que é um problema de otimização que encontra pesos unitários bem comportados e globalmente convexos) mesmo em grandes bases de dados.

Diante do poder da entropia em balancear os grupos de tratados e não tratados, a Tabela 2 demonstra que há um ajustamento exato após a reponderação nas covariadas de participação do ETI nos três momentos da distribuição. Em relação as covariadas, as médias (assim como os outros momentos da distribuição) de alunos que têm pais com ensino superior, a quantidade de moradores no mesmo domicílio e ter frequentado sempre escola pública já são bem ajustadas tanto antes como após o balanceamento, sendo destacada esta última por ter quase 100% de ajustamento entre os grupos de tratados e não tratados (indicando alta homogeneidade da amostra). Quanto as outras variáveis, há desequilíbrios antes do balanceamento entre os grupos, os quais também são ajustados após a amostra ser reponderada

Tabela 2: Balanceamento por entropia.

	Média	Variância	Assimetria	Média	Variância	Assimetria
Antes do Balanceamento						
efeitoDeslocamento	0,060	0,057	3,675	0,012	0,012	8,845
efeitoMigracao	0,133	0,116	2,161	0,198	0,159	1,515
superPai	0,285	0,205	0,952	0,270	0,197	1,035
superMae	0,304	0,212	0,851	0,281	0,202	0,972
moradoresCasa	4,468	2,082	1,058	4,299	1,892	1,450
salarioMinimo	0,380	0,237	0,493	0,705	0,208	-0,899
escPublicaEM	0,985	0,015	-7,922	0,969	0,029	-5,433
Após o Balanceamento						
efeitoDeslocamento	0,060	0,0573	3,675	0,060	0,057	3,675
efeitoMigracao	0,133	0,116	2,161	0,133	0,115	2,160
superPai	0,285	0,205	0,952	0,285	0,204	0,952
superMae	0,304	0,212	0,851	0,304	0,212	0,851
moradoresCasa	4,468	2,082	1,058	4,468	2,145	1,794
salarioMinimo	0,380	0,237	0,493	0,380	0,236	0,493
escPublicaEM	0,985	0,015	-7,922	0,985	0,015	-7,921

Fonte: Elaboração própria dos autores. Nota 1: O método de PSM usado é do vizinhos mais próximos (razão 1:1).

⁶ King e Nielsen (2019) comprovaram por meio de simulações que o MDM é um pareamento alternativo e mais fidedigno as inferências causais do que o PSM.

⁷No estudo de Souza, Ziegelmann e Figueiredo (2018) também há uma discussão sobre as limitações do PSM.

⁸ Um outro estudo que pode auxiliar nos conceitos de entropia e suas vantagens é McMullin e Schonberger (2020).

⁹ Essa forma iterativa é o fato do pesquisador testar diversas equações com diferentes variáveis até encontrar um bom ajuste que melhore os resultados do modelo.

Assim, analisando os ajustes das covariadas a partir dos dois métodos de pré-processamento, observa-se uma consistência nos ajustes para as médias, bem como nos outros momentos. No entanto, como o estudo têm algumas limitações na composição da base de dados (sobretudo nas opções para acrescentar mais covariadas que possam afetar a participação no Programa), como por exemplo, o indivíduo não tem um código de identificação na base do Enem 2017 que permita ligá-lo a outras bases de dados educacionais e/ou até mesmo acompanhá-lo no tempo. Portanto, serão descritas possíveis justificativas, porém pertinentes, que apontam para o alto ajuste entre os grupos de tratados e não tratados mesmo em algumas covariadas antes do pré-processamento:

1. Na estratégia de construção da amostra são realizados alguns filtros específicos — como por exemplo — somente alunos do 3º ano do ensino médio (todos os alunos na amostra estavam cursando e concluíram o 3º ano do ensino médio em 2017) de escolas públicas estaduais (excluindo as que já possuem viés de autosseleção, como as militares/corpo de bombeiros, Educação de Jovens e Adultos, técnicas e profissionalizantes¹⁰) com carga horária maior que 550 minutos (dentro do mínimo estipulado pela Lei Nº 16.287, de 20 de julho de 2017, sendo adotado nessa pesquisa que em 2016 também foi usado esse critério.).
2. Os critérios de escolhas das escolas para se tornarem de tempo integral foram amplificados em 2017¹¹ — porém — já existia um critério de elegibilidade em 2016, o qual a Secretaria de Educação do Estado do Ceará escolhia escolas com 50% ou mais de alunos beneficiários do Programa do Governo Federal Bolsa Família para se tornarem escolas de tempo integral. Portanto, tais critérios não dependiam da escola.
 1. Os principais critérios para elegibilidade do ETI, em 2017, são municípios populosos com jovens em situação de vulnerabilidade social, o que não é tão diferente de jovens que façam parte do Bolsa Família. De qualquer forma, as variáveis que são usadas no balanceamento dos dados contemplam tanto os critérios estabelecidos em 2016, quanto os critérios dissertados na legislação em 2017. Por exemplo, os efeitos de deslocamentos e migração, mesmo que indiretamente, buscam capturar efeitos relacionados ao municípios. Já a quantidade de moradores e ter salário mínimo pode mensurar efeitos inerentes a vulnerabilidade socioeconômica dos estudantes.
 2. A variável que implica se o aluno frequentou todo ensino médio em escolas públicas é ajustada antes e após o pareamento nos dois grupos, indicando enorme grau de homogeneidade. Ou seja, no geral, alunos de baixa renda são os principais frequentadores de escolas públicas no Brasil.
 3. Por fim, há uma quantidade considerável de aluno no grupo de controle para quantidade de alunos no grupo tratados, o que pode está ajudando no alto ajuste.

¹⁰ O estudo de Mariano e Arraes (2018) mostrou a autosseleção das escolas profissionais de ensino médio no Estado do Ceará, da mesma forma que a pesquisa de **Benevides e Soares** (2020) apontou esse viés para as escolas militares do ensino fundamental, também do Estado do Ceará.

¹¹ Em 2016, funcionou como um projeto piloto, onde aquelas escolas com mais de 50% dos alunos beneficiários do Bolsa Família (Programa de Transferência de Renda do Governo Federal para famílias de baixa renda) passariam de tempo parcial para integral. Em 2017, o Programa recebeu novos critérios estabelecidos pela Lei Nº 16.287, de 20 de julho de 2017. Mais informações sobre os critérios podem ser conferidos em <https://www.seduc.ce.gov.br/escola-em-tempo-integral/>.

4.2 Estimações

Nesta seção, as estimações são apresentadas com e sem covariadas. O conjunto de covariadas para controlar possíveis vies que afetam as variáveis de resultado incluem as utilizadas no pre-processamento cujo intuito era mitigar o vies de autosseleção da participação do Programa. Ou seja, as variáveis que afetam a participação do indivíduo no Programa, também podem afetar a variável de resultado. No entanto, as covariadas e algumas outras estatísticas padrões são omitidas nas tabelas com objetivo de melhorar a visualização dos resultados.

A principal estimacão que norteia os impactos das ETI's sobre o desempenho dos estudantes no Enem serão as estimacões reponderadas pela entropia. O OLS serã e a estimacão exploratória, com alto vies, somente para comparacão, assim como PSM, que apresenta várias desvantagens quando se trata de inferência causal, tendo o MDM como alternativa (KING; NIELSEN, 2019; SOUZA; ZIEGELMANN; FIGUEIREDO, 2018). Por outro lado, Hainmueller (2012) demonstrou que a entropia tem melhor desempenho na análise de causalidade, superando todos os métodos tradicionais de pareamento.

Na Tabela 3, pode-se observar que o PSM apresenta resultados positivos significantes estatisticamente, contudo, seus erros padrões são mais elevados do que as estimativas de OLS, porém, mais aproximado das estimativas do MDM. Por outro lado, como esperado, a entropia tem os menores erros padrões e, portanto, aponta estimativas mais eficientes e confiáveis para análise causal do Programa ETI. Os principais resultados da análise são das estimacões com peso calculado por entropia aliado ao uso de covariáveis de características observáveis que buscam mitigar os efeitos do vies no desempenho do alunos nas notas da diversas áreas de conhecimento examinadas pelo Enem 2017.

Em relacão aos coeficientes de impactos, como já explicado anteriormente, a base de dados oferece grupos de tratados e não tratados bastante similares, talvez, seja por isso que os resultados dos métodos tradicionais (PSM e MDM) estejam não tão distantes dos coeficientes de impactos estimados com reponderacão de entropia. No que diz respeito as notas do exame do ensino médio (variáveis de resultados), há um impacto significativo estatisticamente e positivo sobre o aluno de 3º ano que foi beneficiado com o Programa ETI no ano de 2017. O aumento médio desse aluno do ETI em relacão aos alunos que não receberam a intervençã e 17,51 pontos em ciências da Natureza, 42,42 pontos em Ciências Humanas, 37,89 pontos em Linguagens e Códigos, 36,17 pontos em Matemática e 97,95 pontos em Redaçã.

Algumas pesquisas recentes têm usado estratégias na tentativa de avaliaçã de impacto de unidades de ensino que oferecem tempo integral em comparacão com unidades de tempo parcial. No Estado do Ceará, o Programa de Escolas Estaduais de Ensino Profissionalizante (EEEP), que é fundamentado nos mesmos princípios de escolas em tempo integral, contudo, mais direcionado ao mercado de trabalho — Mariano (2016) avaliou esse Programa acompanhando o indivíduo longitudinalmente, utilizando o método PSM, sendo as variáveis de resultados as notas do Enem em 2014. Os achados mostraram que os alunos das EEEP obtiveram maiores médias nas notas do Enem do que alunos de escolas de tempo parcial, sendo cerca de 45,11 pontos em Ciências da Natureza, 54,86 pontos em Ciências Humanas, 58,19 pontos em Linguagens e Códigos, 66,50 pontos em Matemática e 154,32 pontos em Redaçã. Da mesma forma, Araújo, Chein e Pinto (2018) avaliaram a Educaçã Profissional e Tecnológica (EPI) em comparacão com ensino regular, utilizando dados do Enem de 2009 para todo o Brasil. Os resultados relacionados a dependência administrativa estadual apontaram que as escolas EPI tiveram impacto positivo e significativo em todas as notas da prova, sendo Ciências da Natureza (13,82), Ciências Humanas (14,12), Linguagens e Códigos

(19,38), Matemática (13,79) e Redação (32,35).

Tabela 3: Efeito Médio do Tratamento sobre os Tratados (EMTT).

Modelos	Variáveis Dependentes									
	notaCN	notaCN	notaCH	notaCH	notaLC	notaLC	notaMT	notaMT	notaRED	notaRED
OLS	14,11*** (3,70)	13,36*** (4,08)	30,46*** (4,67)	32,45*** (5,07)	28,55*** (4,15)	30,74*** (4,45)	32,75*** (5,21)	31,93*** (5,66)	78,95*** (12,71)	80,06*** (13,77)
PSM	19,35*** (5,02)	17,56*** (5,33)	40,65*** (6,41)	43,94*** (6,59)	33,58*** (5,53)	37,61*** (5,46)	39,16*** (7,61)	37,67*** (7,64)	95,37*** (16,08)	99,07*** (16,57)
MDM	19,10*** (5,00)	21,06*** (5,59)	31,09*** (6,49)	44,10*** (6,96)	27,64*** (5,74)	39,21*** (5,99)	29,47*** (7,43)	32,86*** (8,08)	79,09*** (16,60)	100,73*** (18,05)
Entropia	12,60*** (2,60)	17,51*** (2,83)	29,17*** (3,32)	42,42*** (3,46)	25,79*** (2,91)	37,89*** (2,96)	32,27*** (3,82)	36,17*** (4,01)	72,59*** (8,30)	97,95*** (8,76)
Covariadas	X	C	X	C	X	C	X	C	X	C

Fonte: Elaboração própria dos autores.

(*) significativo a 10%; (**) significativo a 5%; (***) significativo a 1%. Erros Padrões entre parenteses.

Nota 1: As covariadas são características observáveis dos indivíduos que podem afetar tanto a participação no Programa quanto a variáveis de resultados. Nota 2: No PSM e na Entropia, utilizam-se 7 covariadas para o balanceamento. No OLS não há balanceamento.

Nota 3: Nas estimações de resultados são usadas 21 covariadas, incluindo as 7 utilizadas no balanceamento do PSM e Entropia.

Nota 4: O OLS e a Entropia têm 2.050 observações na amostra, enquanto o PSM e MDM tem 524 observações (devido ao descarte).

As pesquisas apresentadas por Mariano (2016) e Araújo, Chein e Pinto (2018) possuem resultados bem distintos dos encontrados na presente pesquisa. Estes podem estar relacionados a diversos fatores, como por exemplo, as estratégias adotadas nas pesquisas subjacentes (como o uso de PSM, já explicado suas desvantagens no estudo de King e Nielsen (2019)), ou o fato dos vários Programas apresentarem desenhos diferentes na sua formação base (o ensino técnico e profissional tem um desenho de política pública diferente daquele modelado para escolas de ensino integral regulares — pois, o primeiro tem um direcionamento mais voltado ao mercado de trabalho). No estudo de Mariano e Arraes (2018) foi introduzida uma discussão sobre os resultados inconclusivos na modalidade de ensino profissional em relação ao mercado de trabalho no mundo, sendo possível também que as altas taxas de desemprego de países como o Brasil possam afetar negativamente os concluintes de ensino profissional a não ingressarem no mercado de trabalho, portanto, eles buscarão competir nos exames de admissão para ingresso no ensino superior, o que tornaria este estudante um forte concorrente a obter notas médias elevadas. Por fim, a avaliação das ETI's utilizam somente alunos que cursaram o 3º ano em 2017, portanto, não há um acompanhamento longitudinal para saber os efeitos no tempo, além de ser improvável que ele tenha cursado 1º e 2º ano do ensino médio em escolas integrais, dado que o Programa foi implantado em 2017.

Outros estudos relacionados ao ensino de tempo integral, porém com diferentes estratégias de pesquisas, como a utilização de variáveis de resultados diferentes, métodos diferentes e/ou abordando ensino fundamental podem ser encontrados para os Estados de São Paulo em Aquino (2011) e Minas Gerais em Soares et al. (2014). No América do Sul, Cerdan-Infantes e Vermeersch (2007) no Uruguai, e Bellei (2009) no Chile.

5 CONSIDERAÇÕES FINAIS

O presente estudo aborda uma literatura pouco explorada, porém, que recebeu bastante espaço nas discussões dos formuladores de política educacionais. O objetivo da pesquisa foi realizar uma avaliação de impacto do Programa ETI sobre as notas do Enem 2017 para o Estado do Ceará, usando métodos de quase experimentos para inferência causal, contribuindo na formação de conteúdo a respeito da efetividade dos resultados.

Em relação aos resultados, todas as disciplinas do Enem 2017 foram influenciadas de forma significativa e positiva pelas ETI's. O foco da análise é apenas nos coeficientes que definem os efeitos de impactos, deixando de lado a análise secundária das variáveis que são responsáveis por controlar o viés. No entanto, é possível observar que retirando o conjunto de variáveis de controle os coeficientes apresentam valores subestimados quando comparado com o método de entropia, cujo os estimadores são os mais eficientes. Ademais, o estudo apresenta uma vantagem no uso de uma metodologia que é superior as abordagens tradicionais (PSM e MDM), sendo esse que esse novo método usa toda a amostra e possui maior eficiência quando são analisados os erros padrões. Portanto, caso os resultados dos parâmetros representem o verdadeiro efeito causal das ETI's sobre as notas do Enem — pode-se afirmar que existem indicativos para uma ampliação do Programa, com o objetivo de reduzir as vulnerabilidades socioeconômicas do Estado e permitir maior desempenho dos beneficiários em exames de admissão ao ensino superior.

Por outro lado, quando se trata de avaliação de impacto causal, os resultados nessa pesquisa devem ser analisados com bastante cautela, tendo em vista que existem diversas limitações, sejam na base de dados, de estratégia e até metodológicas. Primeiro, a base de dados permite somente uma avaliação estática, enquanto seria preferível uma avaliação longitudinal do aluno, que pudesse observar se esses efeitos se mantêm com o tempo. Segundo, a estratégia se limitou ao uso de apenas uma base de dados — o que não possibilitou um maior conjunto de covariadas — isso não foi possível pelo fato dos indivíduos não serem identificados no Enem 2017. Terceiro, a estratégia de encontrar o contrafactual pode não ser a mais adequada — o ideal seria um desenho experimental, onde todos os indivíduos tivessem as mesmas chances de participar das ETI's. Dito isto, pode-se sugerir que além do viés de participação do Programa, o aluno pode também se autosselecionar para fazer a prova do Enem, dependendo da sua motivação em comparação com o aluno da escola de tempo parcial. Portanto, é possível que os coeficientes não representem uma causa e efeito das ETI's nas notas do Enem 2017 — contudo — no cenário mais pessimista possível, os resultados oferecem uma forte correlação nas notas do Enem do aluno associada ao fato dele frequentar 3º ano em ETI.

Outro fato interessante está relacionado a validade externa da pesquisa, a qual pode não obter os mesmos resultados em outros Estados ou usando mecanismos de transmissões diferentes. Além do que, essa pesquisa ainda pode ser complementada usando dados em painel longitudinal, nos quais sejam possíveis uma ligação entre as diversas bases educacionais que possam analisar o Programa ETI ao longo do tempo, tendo em vista a manutenção do Programa e até mesmo redesenhar alguns mecanismos que possam melhorar ainda mais os resultados. Do ponto de vista metodológico, pode ser sugerido também usar uma combinação de entropia com

PSM, buscando examinar uma maior robustez, além de ser importante uma ampliação no estudo para realização de uma avaliação econômica, com intuito de analisar se há retorno no investimento em ensino de tempo integral. Portanto, a pesquisa pretende fornecer conteúdo técnico de qualidade aos formadores de políticas públicas educacionais e mostrar a sociedade civil que suas contribuições tributárias estão sendo revertidas em serviços produtivos que possibilitem bem estar social futuro.

REFERÊNCIAS

AQUINO, J. M. d. Uma ampliação da jornada escolar melhora o desempenho acadêmico dos estudantes? Uma avaliação do programa Escola de Tempo Integral da rede pública do estado de São Paulo. Tese (Doutorado) — Universidade de São Paulo, 2011.

ARAÚJO, A. J. N.; CHEIN, F.; PINTO, C. C. d. X. Ensino profissionalizante, desempenho escolar e inserção produtiva: Uma análise com dados do enem. Instituto de Pesquisa Econômica Aplicada (Ipea), 2018.

BELLEI, C. Does lengthening the school day increase students' academic achievement? results from a natural experiment in chile. *Economics of Education Review*, Elsevier, v. 28, n. 5, p. 629–640, 2009.

BENEVIDES, A. d. A.; SOARES, R. B. Diferencial de desempenho de alunos das escolas militares: o caso das escolas públicas do ceará. *Nova Economia*, SciELO Brasil, v. 30, n. 1, p. 317–343, 2020.

BERNARDONI, D. L. Fortalecimento da função avaliação de políticas públicas: estudo de caso do processo de avaliação da política de emprego e renda da secretaria de estado do trabalho, emprego e promoção social do paraná. 2008.

CAVALIERE, A. M. Quantidade e racionalidade do tempo de escola: debates no brasil e no mundo. *Revista Teias*, v. 3, n. 6, p. 15, 2002.

CERDAN-INFANTES, P.; VERMEERSCH, C. More time is better: An evaluation of the full time school program in Uruguay. [S.l.]: The World Bank, 2007.

COCHRAN, W. G.; RUBIN, D. B. Controlling bias in observational studies: A review. *Sankhyā*: The Indian Journal of Statistics, Series A, JSTOR, p. 417–446, 1973.

DUTRA, P. Educação integral no estado de Pernambuco: uma política pública para o ensino médio. [S.l.]: Editora UFPE, 2014.

FILHO, F. de H. B.; PESSÔA, S. de A. Educação e crescimento: o que a evidência empírica e teórica mostra?

Revista Economia, v. 11, n. 2, p. 265–303, 2010.

GONÇALVES, A. S. Reflexões sobre educação integral e escola de tempo integral. *Cadernos Cenpec| Nova série*, v. 1, n. 2, 2006.

HAINMUELLER, J. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, JSTOR, p. 25–46, 2012.

HIRANO, K.; IMBENS, G. W.; RIDDER, G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, Wiley Online Library, v. 71, n. 4, p. 1161–1189, 2003.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep): Microdados. [S.l.]: Disponível em <<http://inep.gov.br/microdados>>, 2020. Acessado em 03 de novembro de 2020.

- JANNUZZI, P. de M. Considerações sobre o uso, mau uso e abuso dos indicadores sociais na formulação e avaliação de políticas públicas municipais. *Revista de Administração Pública*, v. 36, n. 1, p. 51–72, 2002.
- KING, G.; NIELSEN, R. Why propensity scores should not be used for matching. *Political Analysis*, v. 27, n. 4, p. 435–454, 2019.
- LEE, J.-W.; BARRO, R. J. Schooling quality in a cross-section of countries. *Economica*, Wiley Online Library, v. 68, n. 272, p. 465–488, 2001.
- MARIANO, F. Z. Ensaio sobre o ensino profissionalizante no Ceará. Tese (Doutorado) — Universidade Federal do Ceará, 2016.
- MARIANO, F. Z.; ARRAES, R. A. Trajetória longitudinal dos alunos de escolas profissionalizantes para o mercado de trabalho. *Revista Brasileira de Economia*, SciELO Brasil, v. 72, n. 4, p. 451–477, 2018.
- MCMULLIN, J. L.; SCHONBERGER, B. Entropy-balanced accruals. *Review of Accounting Studies*, Springer, p. 1–36, 2020.
- MILITÃO, S. C. N.; KIILL, A. N. A escola de tempo integral no Brasil: trajetória histórica e legal. *Colloquium Humanarum*, v. 12, n. Especial, 2015.
- ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, Oxford University Press, v. 70, n. 1, p. 41–55, 1983.
- SIMÕES, L. G.; CIRINO, J. F.; CASSUCE, F. C. da C. O impacto da educação no rendimento do trabalho: estudo sobre a quantidade e a qualidade. *Revista de Ciências Humanas*, n. 2, 2014.
- SOARES, T. M. et al. Escola de tempo integral: resultados do projeto na proficiência dos alunos do ensino fundamental das escolas públicas da rede estadual de Minas Gerais. *Ensaio: Avaliação e Políticas Públicas em Educação*, SciELO Brasil, v. 22, n. 82, p. 111–130, 2014.
- SOUZA, W. P. S. de F.; ZIEGELMANN, F. A.; FIGUEIREDO, E. A. de. As condições de saúde afetam os rendimentos do trabalho? evidências para o mercado de trabalho no Brasil. *Economia Aplicada*, v. 22, n. 4, p. 113–150, 2018.
- WOESSMANN, L. The economic case for education. *Education Economics*, Taylor & Francis, v. 24, n. 1, p. 3–32, 2016.

IMPACTO DESIGUAL DO FINANCIAMENTO DA EDUCAÇÃO BASEADO EM RESULTADOS

Pedro Alexandre Santos Veloso*
Rafael Barros Barbosa**

RESUMO

O Ceará foi o primeiro estado a introduzir uma política de redistribuição de transferências tributárias baseadas na performance educacional agregada dos municípios, alcançando resultados educacionais relevantes. O presente artigo analisa se a Lei da Cota Parte de ICMS implementada em 2009 muda a composição dos gastos municipais, em especial o aumento de gastos em educação, bem como o impacto dessa política sobre os resultados educacionais das escolas municipais. Os resultados mostram que receber mais recursos da Lei da Cota Parte não produz um elevado gasto em educação, mas sim um aumento no gasto total não educacional. Especificamente, para cada real recebido pelos municípios devido a Lei da Cota Parte, foram gastos R\$ 0,45 centavos em educação (R\$ 0,23 no Ensino Fundamental) e R\$ 1,96 em gastos totais per capita. As estimativas do segundo resultado, sugerem que o modelo de repartição de recursos baseados na performance municipal tem impacto relevante apenas sobre as escolas que possuíam elevada performance antes da implementação da Lei da Cota Parte. Foi evidenciado uma possível discriminação no gasto público em educação para as escolas de melhor performance antes da introdução da Lei da Cota Parte. A implicação direta é que tal política aumentou a desigualdade educacional entre as escolas.

Palavras-Chaves: Financiamento da educação, transferências intergovernamentais, performance educacional.

ABSTRACT

The state of Ceará was the first state to introduce a policy of redistributing tax transfers based on the aggregate educational performance of the municipalities, achieving relevant educational results. This article examines whether the ICMS Law implemented in 2009 changes the composition of municipal spending, in particular the increase in spending on education, as well as the impact of this policy on the educational results of municipal schools. The results show that receiving more resources from the ICMS Law does not produce a high expenditure on education, but an increase in total non-educational expenditure. Specifically, for each real received by the municipalities due to the Quota Part Law, R\$ 0.45 cents was spent on education (R\$ 0.23 in Elementary Education) and R\$ 1.96 on total expenses per capita. Estimates of the second result suggest that the model of allocation of resources based on municipal performance has a relevant impact only on schools that had high performance before the implementation of the Law of the Share Quota. There was evidence of possible discrimination in public spending on education for the best performing schools before the introduction of the Lei da Share Quota. The direct implication is that such a policy has increased educational inequality between schools.

Keywords: Educational financing, intergovernmental transfers, educational performance.

Área Temática: Área 4 - Setor Público

JEL: I22, I28, H70

* CAEN- UFC, Brasil. E-mail: p.veloso@caen.ufc.br. Tel:(85)988430-8046.

** Departamento de Economia Aplicada - UFC, Brasil. E-mail: rafael.barbosa@ufc.br. Tel: (85) 99930-1938.

1 INTRODUÇÃO

Um dos fatores mais importantes da oferta de educação pública corresponde a forma de financiamento. Evidências recentes apontam que a quantidade de recursos despendidos importa para os resultados educacionais (JACKSON, JOHNSON e PERSICO (2016); JACKSON, WIGGER e XIONG (2020); HADDAD, FREGUGLIA e GOMES (2017)). No entanto, existe forte heterogeneidade na relação entre gasto público e educação, especialmente no Brasil, sugerindo que a forma como tal gasto é implementado também é importante para os resultados. Diante de incentivos difusos, a oferta da educação pública com qualidade pode não ser alcançada.

Uma política que tem ganhado destaque recentemente é a distribuição de recursos públicos educacionais baseado na performance ou no mérito dos responsáveis pela educação. Basicamente, este tipo de política consiste em distribuir mais recursos públicos para entidades ou indivíduos que obtêm melhores resultados educacionais. Diferentes modelos já foram aplicados e testados¹ ao redor do mundo e seus resultados mostram-se promissores. Os modelos mais tradicionais redistribuem recursos ao nível da escola (FIGLIO e ROUSE (2006)), do professor (JACOB (2005); REBACK (2008); FERNANDES e FERRAZ (2014); MBITI et. al. (2019)) ou dos pais (FIGLIO e LUCAS (2004)). A principal vantagem da adoção deste tipo de política é a criação de incentivos direcionados a melhoria da educação. Por outro lado, críticos defendem que tais incentivos podem substituir a motivação intrínseca da educação e aumentar a desigualdade entre as entidades ou indivíduos (FIGLIO E GETZLER (2006); CULLEN e REBACK (2006); NEAL e SCHANZENBACH (2010)).

Uma novidade introduzida pelo estado do Ceará em 2009 foi a redistribuição de transferências tributárias baseadas na performance educacional agregadas dos municípios. Diferente de outras estratégias que focam em desempenho da escola, do professor ou diretamente dos estudantes, a Lei nº 14.023 de 2007 (doravante Lei da Cota Parte ou LCP) definiu critérios para distribuição de recursos da cota parte do ICMS de acordo com o desempenho educacional dos municípios no Ensino Fundamental I. Este desempenho é mensurado de forma agregada entre todas as escolas do município. Dos 25% de recursos devidos aos municípios derivados da arrecadação estadual com o ICMS, 18% são distribuídos de acordo com tais critérios. Os demais 7% são distribuídos de acordo com a qualidade da saúde e do meio ambiente municipal. As evidências empíricas iniciais apontam que tal política aumentou a performance dos estudantes². A figura 1, disponível no Apêndice, resume parte destas evidências. Analisando a performance média dos municípios no IDEB (Índice de Desenvolvimento da Educação Básica), percebe-se que o Ceará apresentou uma evolução significativa na aprendizagem no 5º ano do Ensino Fundamental, mesmo estando em situação socioeconômica adversa.

Duas características interessantes deste modelo de redistribuição de recursos são: 1. Os recursos são fungíveis, isto é, os prefeitos que recebem os recursos não são obrigados a gastá-los em educação, podendo optar por dar outros destinos; 2. Os resultados são mensurados de forma agregada ao nível do município. Este último ponto permite que municípios possam

¹ Para uma revisão recente desta literatura, ver LEE e MEDINA (2019).

² São exemplos desta literatura: Brandão (2014); Carneiro e Irffi (2018); Petterini e Irffi (2013); Shirasu, Irffi e Petterini (2013). Uma limitação destes trabalhos decorre da ausência de controles para políticas que tenham sido implementadas contemporaneamente a Lei da Cota Parte. No caso do Ceará um programa importante que pode confundir os resultados é o Programa de Alfabetização na Idade Certa (PAIC). Evidência recente (MURALIDHARAN et al (2019)) mostra que a combinação de políticas de gasto por performance e incentivos a sua adequada implementação podem explicar a maior parte dos resultados.

ter uma performance média elevada, mesmo se algumas escolas não melhorarem seus resultados educacionais.

A despeito das evidências favoráveis a eficiência da política, algumas questões ainda não foram devidamente endereçadas. Não há evidências, por exemplo, do impacto da LCP sobre a composição dos gastos municipais, i.e., municípios que receberam mais transferências, aumentam o gasto em educação? Este ponto é importante para entender se a LCP induz uma busca por melhorias na qualidade educacional por meio do aumento de gastos. Adicionalmente, não há evidências dos impactos desta política sobre a desigualdade entre as escolas em diferentes níveis de proficiência. Isto é, a política de redistribuição da Lei da Cota Parte aumentou ou diminuiu a desigualdade escolar de acordo com a performance prévia? Além disso, prefeitos podem alocar recursos de forma discriminada para escolas que consigam gerar melhores resultados, possibilitando manter um nível de performance agregado estável ao longo do tempo. Responder essa última pergunta é importante pois ajuda a entender como os prefeitos optaram por alocar seus recursos nas escolas em resposta ao incentivo gerado pela LCP. Os resultados ideais dessa política seriam que os incentivos da LCP aumentassem o gasto em educação ao nível municipal e buscassem melhorar o desempenho das escolas com piores performances, reduzindo a desigualdade intramunicipal.

O presente artigo tem o objetivo responder tais perguntas. Primeiro, é investigado se o fato de um município receber mais recursos com a implantação da Lei da Cota Parte em 2009 implica em maiores gastos em educação. Por meio de uma estratégia empírica de diferença em diferença, verifica-se que receber mais recursos da Lei da Cota Parte não produz um elevado gasto em educação, mas sim um aumento no gasto total não educacional. Especificamente, para cada real recebido pelos municípios devido a Lei da Cota Parte, foram gastos R\$ 0,45 centavos em educação (R\$ 0,23 no Ensino Fundamental) e R\$ 1,96 em gastos totais per capita.

Segundo, explorando a variabilidade intermunicipal antes da implementação da Lei da Cota Parte, constata-se que escolas com baixa performance em 2007 não aumentaram seus resultados educacionais em municípios que se beneficiaram com a LCP quando comparados com municípios que perderam recursos com a política. Por outro lado, escolas com melhores performance em 2007 aumentaram seus resultados educacionais em municípios que se beneficiaram com a Lei da Cota Parte. Em resumo, as estimativas sugerem que a LCP tem impacto relevante apenas sobre as escolas que possuíam elevada performance antes da implementação da Lei da Cota Parte. A implicação direta é que tal política aumentou a desigualdade educacional entre as escolas. Os resultados são robustos à inclusão de características pré-determinadas das escolas, excluindo parcialmente o fato de tais escolas terem fatores prévios que contribuem no desempenho futuro em testes padronizados.

Tomados em conjunto, os resultados sugerem que a política de redistribuição baseada no mérito educacional dos municípios não induziu a um gasto maior em educação, além de aumentar desigualdade educacional ao nível da escola. Quanto ao primeiro ponto, a Lei da Cota Parte ao permitir que o gasto seja desvinculado do gasto em educação, este pode ser um resultado desejável pelos legisladores que a elaboraram. Quanto ao segundo ponto, a principal explicação é a presença de uma alocação de recursos intramunicipalmente direcionada para as escolas que têm maior potencial de gerar bons resultados nos testes padronizados.

Para verificar esta hipótese de alocação diferenciada de recursos foi estimado um modelo cujas variáveis de resultado são: Horas Aula Média Por Dia e Média de Alunos por Turma. Ambas as variáveis estão associadas a qualidade da oferta de educação pública e também a maiores gastos em educação. Os resultados sugerem que as melhores escolas em municípios beneficiados com a LCP reduziram significativamente o tamanho médio das turmas. Esse mesmo resultado não foi encontrado para as escolas com mais baixas performances prévias mas que também estejam em município beneficiados pela LCP. Não foi verificado diferença

significativa em relação ao número médio de horas diárias, embora a magnitude das estimativas suporte a validade da hipótese de alocação diferenciada entre as escolas.

Este trabalho contribui para diferentes áreas da pesquisa em economia da educação. Primeiro, para uma ampla literatura sobre a importância dos gastos públicos para a qualidade da educação (HANUSHECK (2005); MANUELLI e SESHADRI (2014); HADDAD, FREGUGLIA e GOMES (2017)). Segundo, buscando dar incentivos adequados, muitas políticas são desenhadas de forma a premiar a performance. Pesquisas recentes têm sido realizadas sobre este tema e este artigo contribui ao analisar como agentes incentivados utilizam tais recursos e se eles potencialmente geram alocações diferenciadas de recursos. Por fim, o financiamento da educação pública é um desafio em países com poucos recursos como o Brasil. Dessa forma, este artigo contribui para entender quais os resultados de diferentes tipos de políticas de repartição de tributos intragovernamentais. Este último ponto é importante tendo em vista a aprovação do novo FUNDEB que possui uma parte inspirada no modelo cearense de repartição de recursos. O que o presente artigo mostra é que o desenho dos incentivos importa para obter determinados resultados relevantes para políticas públicas.

Além desta introdução, este artigo está subdividido em mais 4 seções. A seção seguinte apresenta em detalhes a Lei da Cota Parte. A seção 3 discute a base de dados utilizada neste trabalho e as estratégias empíricas adotadas. A seção 4 reporta e discute os resultados principais. Por fim, a seção 5 comenta as conclusões gerais do trabalho.

2 POLÍTICA DE TRANSFERÊNCIAS BASEADA EM RESULTADOS

O estado do Ceará fornece um exemplo relevante de como superar as condições socioeconômicas adversas para melhorar substancialmente os resultados da educação. Por meio de políticas de financiamento baseado em resultados (FBR) como parte de um programa mais abrangente de reforma educacional, o Ceará deu início a significativas reformas educacionais que permitiram melhorar os níveis de aprendizagem desde da alfabetização até dos alunos do ensino fundamental (Loureiro *et. al.* 2020).

O sucesso cearense é baseado em alguns pilares interdependentes, nos quais se pode citar: 1. Incentivos financeiros para que os municípios atinjam as metas educacionais estabelecidas (Lei Cota Parte); 2. Assistência técnica aos municípios com dificuldades para melhorar o aprendizado, com ênfase na alfabetização na idade certa para todos (PAIC³); 3. Estabelecimento de um sistema de monitoramento e avaliação sólida e confiável que mede continuamente os principais resultados da educação, incluindo a aprendizagem dos alunos (SPAECE).

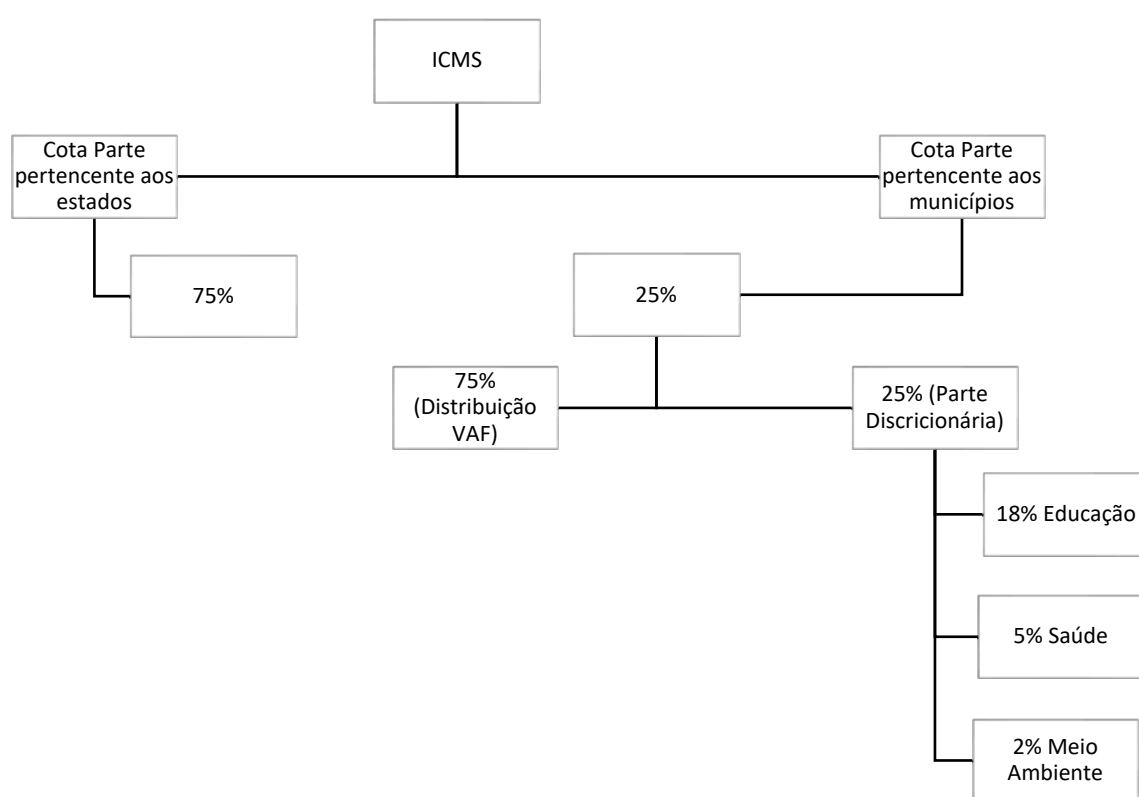
A Constituição Federal confere aos estados a discricionariedade para definir as regras de repartição da cota parte do ICMS com seus municípios. Esse valor refere-se a 25% do total arrecadado pelo estado com o tributo. A novidade implementada pelo estado do Ceará foi utilizar tal discricionariedade para induzir governos municipais na busca de melhoria de resultados sociais, em particular na educação. Esta possibilidade ainda é pouco explorada no Brasil. A figura 2, no Apêndice, extraída de Simões e Araújo (2019), mostra como os estados têm adotado estratégias diversas para distribuir os recursos do ICMS. A maior parte dos estados adotam critérios de repartição baseados no tamanho populacional dos municípios ou dos estudantes. Exceção aos estados do Ceará, Pernambuco, Amapá, Minas Gerais, Bahia e Espírito Santo que adotam critérios baseados na qualidade da prestação do serviço público. No entanto, o Ceará se destaca neste conjunto de estados por ter o maior percentual da cota parte do ICMS vinculada a essa forma de distribuição.

³ Programa de Alfabetização na Idade Certa.

Perante um cenário de reformas estruturais definidas pelo Governo Nacional em 1996, por meio da Lei de Diretrizes e Bases da Educação Nacional (LDB)⁴, o estado do Ceará inovou ao colocar a educação no centro da distribuição da cota parte, a partir da Lei nº 14.023 em 2007, regulamentada em 2008 e passando a valer em 2009. Assim, a parcela discricionária da cota parte (25% do total arrecadado com o ICMS) passou a ser de 72% destinado para educação, 20% para saúde e 8% para meio ambiente. Esta nova forma de distribuição modifica critérios que vigoravam anteriormente desde 1996⁵. A figura 3 esquematiza a nova distribuição de ICMS para os municípios cearenses.

Ao considerar indicadores claros, o mecanismo baseado em resultados têm por finalidade a transparência e geração de incentivos para os municípios melhorarem seus resultados. Para tanto, o governo estadual criou índices de qualidade da educação que consideram o nível e as melhorias na alfabetização dos alunos da 2ª série, o desempenho dos alunos da 5ª série em leitura e matemática e as taxas médias de aprovação da 1ª para a 5ª série⁶.

Figura 3: Estrutura de repasse de ICMS



Fonte: Elaboração Própria, com base na Lei nº 14.023 de 2007

A partir de 2012, quando a computação da formação do índice educacional (IQE) foi alterado pelo decreto nº 30.796 de dezembro de 2011, o repasse 18% da cota-parte do ICMS aos municípios, ficou atrelado a soma ponderada dos seguintes componentes: (i) 50% dos recursos são distribuídos conforme o índice de qualidade da alfabetização (IQA), calculado a partir dos resultados no exame da alfabetização dos estudantes do 2º ano do ensino fundamental; (ii) 45% em função do índice de qualidade do fundamental (IQF), medido em relação

⁴ Lei nº 9.394 de 20 de dezembro de 1996.

⁵ A Lei nº 12.612 de 1996, estabeleceu que a distribuição da cota parte do ICMS dos municípios deveria observar os seguintes critérios: 75% em conformidade ao valor adicionado fiscal; 12,5% relativo a proporção dos gastos em educação sobre a receita municipal; 7,5% equitativo a todos os municípios; e 5% proporcionais à população de cada município.

⁶ Para uma revisão da implementação desta política, ver Simões e Araújo (2019); Brandão (2014).

desempenho dos alunos do 5^a ano nas provas de português e matemática; e (iii) 5% em razão da taxa média de aprovação dos estudantes do 1^o ao 5^o ano.

O IQA leva em consideração o resultado médio do teste de alfabetização do município multiplicado pelo fator de universalização e pelo percentual de alunos avaliados, gerando uma média transformada que é comparada em nível e em evolução entre anos considerados. O fator de universalização considera o percentual de alunos avaliados como não alfabetizados, alfabetização incompleta e em nível desejado. A redução dos dois primeiros grupos, bem como o aumento do terceiro, melhora a média do município. O percentual de alunos avaliados é uma medida de dispersão no qual o objetivo é evitar que os gestores incentivem somente os melhores alunos a realizarem o exame.

O IQF também corresponde a uma média transformada que é comparada em nível e em evolução entre anos considerados e segue uma lógica de cálculo semelhante ao IQA. São considerados separadamente os resultados médios obtidos pelos alunos nas avaliações de língua portuguesa e matemática multiplicado pelo percentual de alunos avaliados no ano em questão e pelo fator de universalização do aprendizado, que classifica os resultados entre muitos críticos e adequado.

3 BASE DE DADOS E ESTRATÉGIA EMPÍRICA

3.1 Base de dados

A base de dados é dividida em duas partes. Para avaliar o efeito da Lei da Cota de Parte utilizou-se informações da base de dados Finanças Municipais – FINBRA, disponibilizados pela Secretaria do Tesouro Nacional – STN. Foram coletados desta base de dados as seguintes informações municipais: tamanho da população, gasto total, gasto com educação, gasto com ensino fundamental, PIB, transferências de ICMS, fundo de manutenção e desenvolvimento da educação básica – FUNDEB e fundo de participação dos municípios – FPM. O período considerado inicia-se em 2004 e finaliza em 2017. Todas as informações monetárias foram dessazonalizadas utilizando o IPCA-Fortaleza, disponível no Instituto Brasileiro de Geografia e Estatística – IBGE. Os coeficientes de rateio da cota-parte do ICMS e os índices de educação, saúde e meio ambiente no período de 2009 a 2017 foram obtidos junto ao IPECEDATA.

Para avaliar o impacto das mudanças na regra de distribuição do ICMS sobre a proficiência de alfabetização dos alunos do 2^o ano do Ensino Fundamental do Ceará foram considerados os microdados do SPAECE-Alfa disponibilizados pela SEDUC/CE durante os anos de 2007 a 2015, além dos dados do Censo Escolar divulgados pelo Instituto de Pesquisas Educacionais Anísio Teixeira – INEP, sobre as características das escolas e das turmas. As variáveis selecionadas incluem algumas características dos professores, tais como, idade, cor/raça, escolaridade. De forma complementar, utilizou-se indicadores de fluxo escolar como aprovação, reprovação, abandono e distorção idade-série, disponibilizados pelo INEP. A descrição de todas as variáveis utilizadas encontra-se disponível no quadro 1 no Apêndice.

Os municípios de Abaiara, Altaneira, Cariré, Groaíras, Ibaretama, Palmácia, São Benedito, São Luís do Curú, Uruburetama foram excluídos da amostra por não terem dados suficientes das escolas para que a estratégia empírica fosse implementada. Vale ressaltar que o SPAECE-Alfa utiliza a Teoria de Resposta ao Item (TRI) que permite a comparação entre unidades amostradas em diferentes níveis de ensino ao longo do tempo.

3.2 Estratégia Empírica

3.2.1 Estratégia Empírica para impacto da LCP sobre os gastos públicos municipais

A primeira parte deste artigo busca responder a seguinte pergunta: qual o impacto de se receber mais recursos da Lei da Cota Parte sobre os gastos educacionais ou sobre os gastos totais? Para responder a essa questão será adotada uma estratégia empírica de diferença em diferença.

A equação de interesse é definida por:

$$\Delta y_{gmt} = \sum_{t=2004}^{2007} \beta_{1t} \Delta CP_{mt} \times I(m = Q_1) + \sum_{t=2009}^{2017} \beta_{2t} \Delta CP_{mt} \times I(m = Q_1) \times I(\text{Após}) + \tau_m + \tau_t + \gamma_{mt} + \varepsilon_{gmt} \quad (1)$$

Em que: Δy_{mt} é a diferença entre um tipo de gasto $g = \{\text{Gasto Total, Gasto em Educação, Gasto no Ensino Fundamental}\}$ em relação ao gasto do ano de 2008 per capita. O ano de 2008 foi escolhido como referência por ser o último ano antes da implementação da Lei da Cota Parte. Adicionalmente foi considerado como referência o tamanho da população de 2008, para evitar que variações populacionais afetem os resultados. A definição de Δy_{mt} é a que segue:

$$\Delta y_{mt} = \frac{(G_{gmt} - G_{gm2008})}{(Pop_{m2008})} \quad (2)$$

Por sua vez, ΔCP_{mt} é a diferença entre a receita da Cota Parte do ICMS em relação a quanto o município recebia em 2008 em termos per capita. Ou seja,

$$\Delta CP_{mt} = \frac{(CP_{mt} - G_{m2008})}{(Pop_{m2008})} \quad (3)$$

Como antes de 2009, ano da implementação da Lei da Cota Parte, o recurso da Cota Parte do ICMS era dividido de acordo com outros critérios, tal variável ΔCP_{mt} mensura o quanto cada município passou a receber a mais ou a menos a partir de 2008. A LCP fez com que municípios com melhores performances educacionais fossem beneficiados com a aprovação da Lei da Cota Parte em relação a municípios que não tinham boas performances educacionais.

Adicionalmente, τ_t e τ_m são efeitos fixos anuais e municipais cuja função é capturar variações idiossincráticas em determinado ano (choques econômicos, secas, etc.) ou município. Gastos Públicos em geral estão fortemente associados ao ciclo econômico dos respectivos municípios. Municípios com maior variabilidade no ciclo econômico podem ter gastos menores do que municípios com menor variação cíclica (ZIDAR, 2019). Assim, para controlar tais efeitos de ciclos são introduzidos os efeitos fixos do ciclo econômico de cada município variando no tempo⁷, γ_{mt} . Na análise de robustez é verificado se os resultados permanecem os

⁷ Para estimar tais efeitos fixos do ciclo econômico foi computado a variância da taxa de crescimento anual do PIB real de cada município entre os anos de 1999 e 2017. Posteriormente, tais variâncias foram clusterizadas em quatro grupos caracterizando diferentes ciclos econômicos para cada município. Esse ciclo foram interagidos com a variável ano e adicionado como efeitos fixos na equação (1).

mesmos ao substituir γ_{mt} por efeitos fixos municipais variantes no tempo. O período de análise considerado inicia-se em 2004 e finaliza em 2017.

A variável $I(m = Q_1)$ designa os municípios que serão considerados tratados. Esta é uma variável indicadora que atribui valor 1 para os municípios que tinham melhor desempenho educacional antes da implementação da Lei da Cota Parte em 2009. Estes municípios são aqueles que possuem maior probabilidade de aumentarem suas parcelas na redistribuição da cota parte do ICMS. Além disso, $I(m = Q_1)$ atribui valor zero para todos os demais municípios. Para definir quais municípios possuem melhor desempenho educacional antes da implementação da Lei da Cota Parte foi utilizado o índice de qualidade da educação (IQE) para o ano de 2009 (IQE_{m2009}). O IQE_{m2009} mensura a performance agregada dos municípios para os anos de 2007 e 2008. Os municípios foram ranqueados segundo o IQE_{m2009} e o 30º percentil (os 30% municípios mais bem ranqueados) foram considerados tratados. Os 70% restantes foram considerados controles. Erros-padrão foram estimados ao nível do município.

Para tal estratégia seja válida o IQE_{m2009} , uma variável pré-determinada, deve satisfazer duas condições. Primeira, deve ser correlacionado com ΔCP_{mt} a partir de 2009. Essa condição é plausível, pois o IQE_{m2009} mensura a performance da educação municipal para os anos de 2007 e 2008 e com base neste indicador os municípios passaram a receber mais ou menos recursos da cota parte. Segundo, o IQE_{m2009} não deve ser influenciado pelos municípios visando receber mais recursos quando a Lei da Cota Parte for implementada. Acredita-se que tal hipótese também seja plausível por dois motivos: 1. A lei 14.023 que estabelece a nova repartição da cota parte foi publicada no final do ano de 2007 (17/12/2007), com isso, o gasto em educação em 2007 não foi afetado pela LCP; 2. Municípios podem ter aumentado o gasto em 2008 para obter mais recursos decorrentes da Lei da Cota Parte no ano seguinte, porém, acredita-se que um ano não é suficiente para que modificações substanciais ocorram na performance educacional municipal. Assumindo a validade de tais hipóteses, o IQE_{m2009} mensura o quanto cada município irá receber a partir de 2009 e não está associado com mais ou menos gasto em educação no período anterior à promulgação da Lei da Cota Parte.

Tais hipóteses poderão ser verificadas na estimação dos parâmetros β_{1t} que representam a diferença de gasto entre tratados e controle antes da Lei da Cota Parte. Caso a hipótese de tendências paralelas seja válida, tais parâmetros não serão significativos. O conjunto de parâmetros de interesse são β_{2t} que indicam as diferenças de gasto entre os municípios tratados e controle após a implementação da Lei da Cota Parte. O período após a implementação da Lei da Cota Parte é designado pela variável indicadora $I(\text{Após})$.

3.2.2 Estratégia Empírica para mensurar efeito da LCP sobre desigualdade educacional

Este artigo também busca entender se o impacto da promulgação da Lei da Cota Parte afetou de forma diferenciada o desempenho das escolas no 2º ano do Ensino Fundamental. Para tanto, será explorada a variabilidade na qualidade das escolas antes da introdução da Lei da Cota Parte, uma abordagem semelhante à usada por Cilliers et al (2020), que usa um modelo de diferença em diferença. Em síntese, tal abordagem explora a variabilidade na qualidade das escolas intermunicipal mente, mensurada antes da introdução da Lei da Cota Parte.

Em 2007 foi aplicado o teste do SPAECE-Alfa para todos os estudantes do 2º ano do Ensino Fundamental. Com base nesse teste, as escolas em cada município $m = 1, \dots, 184$, foram divididas em quatro categorias de acordo com sua performance média: baixo desempenho, médio desempenho, médio alto desempenho e alto desempenho. Para facilitar a exposição tais categorias serão indexadas em $k = 1, 2, 3$ e 4 , respectivamente.

O objetivo aqui é verificar se houve diferenças significativas entre as escolas de cada categoria que estão em municípios que em 2009 serão beneficiados ou não com a introdução da Lei da Cota Parte. Novamente, os municípios beneficiados foram obtidos pelo ranqueamento do IQE_{m2009} . Os municípios desta vez estão divididos em três grupos: Municípios que se beneficiaram com a Lei da Cota Parte (doravante Beneficiados), municípios neutros que ou tiveram pouco benefício ou pouco prejuízo (doravante Neutros) e municípios que apresentaram perdas em termos de recursos da cota parte com a nova forma de repartição (doravante Prejudicados). Essa separação levou em consideração os quartis em que os municípios se localizam⁸ de acordo com o IQE_{m2009} . Defina a variável indicadora $I(CP_m) = 1$ se o município m está no grupo dos Beneficiados e $I(CP_m) = 0$ se o município m pertence ao grupo do Prejudicados.

Considere y_{qmt} a performance média das escolas no SPAECE-Alfa do município m , no tempo t e nas categorias $Q = 1, 2, 3, 4$. A equação empírica a ser estimada para o segundo exercício é:

$$y_{qmt} = \beta_0 + \beta_k I(CP_m) + \tau_t + \theta_{mt} + u_{kmt} \text{ para as categorias } q = 1, 2, 3 \text{ e } 4 \quad (4)$$

O parâmetro de interesse é β_k que mensura o efeito de dado município estar no grupo dos Beneficiados em comparação ao grupo dos Prejudicados ($I(CP_m)$) para cada categoria k das escolas. Assim, será comparado se existe efeito (positivo ou negativo) em ter alto ou baixo desempenho anterior à Lei da Cota Parte e estar em um município que foi Prejudicado ou Beneficiado pela Lei em termos de ganhos de recursos intragovernamentais. τ_t representa efeitos fixos temporais e θ_{mt} os efeitos fixos municipais variantes no tempo. Erros padrões foram estimados ao nível da escola.

A estimativa do parâmetro β_k pode ser considerada causal por dois motivos: Primeiro, a introdução da Lei da Cota Parte no final do ano de 2007 potencialmente não afetou a performance das escolas neste mesmo ano, pois o exame do SPAECE-Alfa foi realizado antes da promulgação da Lei da Cota Parte. Segundo, a introdução da Lei da Cota Parte pode ser considerada exógena temporalmente à performance educacional das escolas em cada município, isto é, antes da introdução da Lei da Cota Parte escolas tinham diferenças nas performances educacionais intramunicipalmente não correlacionadas com os recursos intermunicipais futuros.

Uma limitação desta abordagem representada pela equação (2) decorre de a possibilidade das escolas entre as categorias não serem comparáveis entre si antes da introdução da LCP. Para contornar esse problema são adotadas duas estratégias. Primeiro, é acrescentado um grande conjunto de variáveis de controle pré-determinadas relativas ao ano de 2007. Isso permite controlar para características observáveis entre as escolas mensuradas antes da introdução da LCP. Segundo, é realizado um pareamento por Entropia (Hainmueller, 2012) utilizando variáveis pré-determinadas para o ano de 2007 ao nível da escola. O pareamento permite atribuir pesos a semelhança entre as escolas, possibilitando uma comparação mais adequada entre tratados e controle. Ambos procedimentos tem o objetivo de permitir que tais escolas sejam comparáveis em cada categoria k .

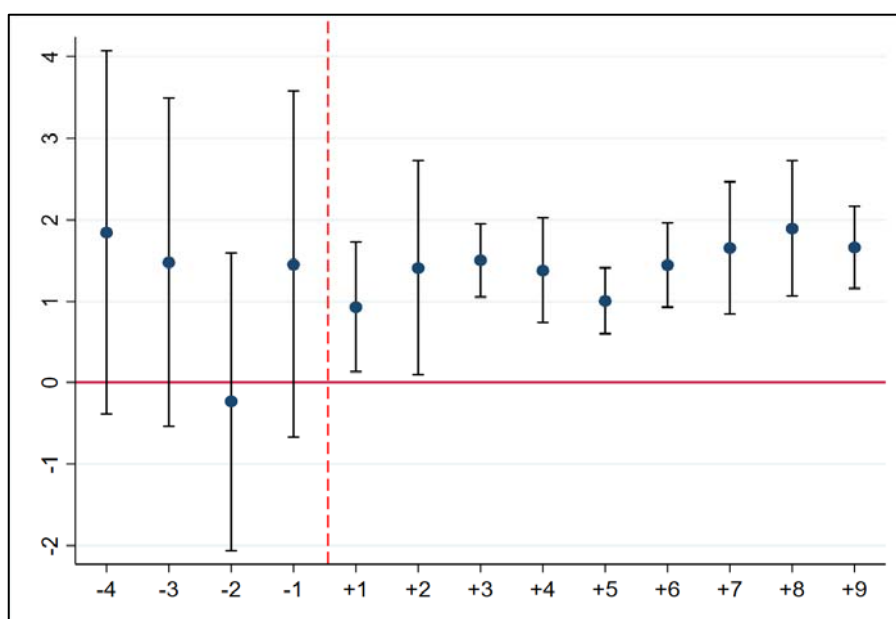
⁸ Notem que os municípios considerados Beneficiados estão no 33º maior centil, os municípios Neutros estão entre o 34º e o 65º centil. Por fim, os municípios Prejudicados estão no 66º até o 100º centil.

4 RESULTADOS

4.1 Resultado para gastos

As figuras 4 e 5 apresentam os resultados gráficos da estimação da equação (1). Especificamente, a figura 4 refere-se ao gasto total, e a figura 5 refere-se ao gasto em educação (Painel A), incluindo todos os tipos de oferta educacional que o município realize⁹, e o gasto especificamente no Ensino Fundamental (Painel B). Importante notar que o Ensino Fundamental é a etapa em que a performance municipal agregada é mensurada para computar o quanto os municípios irão receber em relação a cota parte.

Figura 4: Efeito sobre o Gasto Total da Lei da Cota Parte



Nota: As estimativas anteriores a implementação da LCP (demarcada pela linha pontilhada em vermelho) informam que não havia diferença entre os gastos totais do municípios com elevada performance educacional em comparação aos demais municípios. Já os resultados posterior a implementação indicam que gasto total municipal foi maior que a quantidade recebida inicialmente pela LCP, percebe-se que tal elevação é persistente no tempo, sugerindo que a LCP produziu modificações de longo prazo na quantidade recursos despendidos.

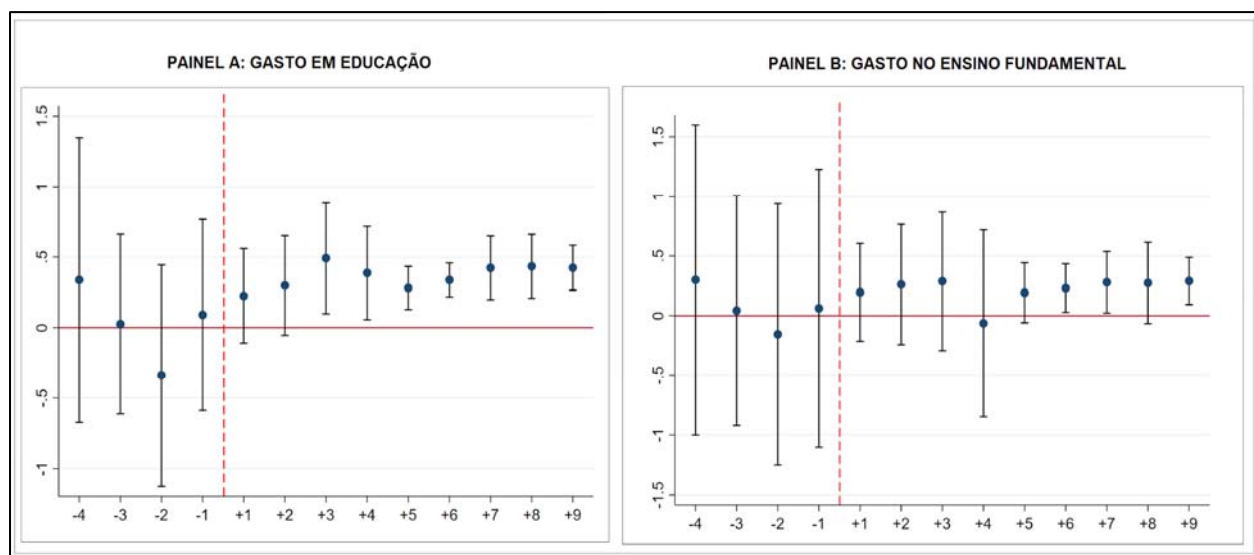
Observe, inicialmente, que as estimativas anteriores à implementação da Lei da Cota Parte (demarcada pela linha pontilhada em vermelho) são não significativas. Isso é uma evidência favorável à hipótese de tendências paralelas requerida para a validade da estratégia da diferença em diferença. As estimativas informam que não havia diferença entre os gastos totais do municípios com elevada performance educacional em comparação aos demais municípios antes da introdução da LCP. Ou seja, os recursos da cota parte anteriores a 2008 não explicam as diferenças entre os padrões de gasto dos municípios.

Analisando as estimativas posteriores à implementação do programa, percebe-se que o gasto total municipal aumentou em aproximadamente R\$1,96 centavos a cada real recebido pelo município decorrente da distribuição da cota parte do ICMS. Essa evidência sugere que a Lei da Cota Parte produziu um aumento relevante dos gastos totais, maior do que a quantidade monetária recebida inicialmente.

⁹ Nesta categoria incluem-se gastos na Educação de Jovens e Adultos (EJA) e outros gastos associados a programas realizados pelo município.

Este fenômeno é conhecido na literatura como *Efeito Flypaper* segundo o qual os entes públicos tendem a gastar mais do que os recursos que recebem de transferências intergovernamentais¹⁰. No caso cearense, percebe-se que os municípios quase dobraram seus gastos totais frente a um aumento de recursos da LCP. Além disso, e talvez mais importante, tal elevação no gasto total é persistente no tempo, sugerindo que a LCP produziu modificações de longo prazo na quantidade de recursos totais despendidos pelos municípios.

Figura 5: Efeito da Lei da Cota Parte sobre o Gasto em Educação e no Ensino Fundamental



Nota: O painel A mostra os gastos em educação. Percebe-se que o gasto aumentou em decorrência da LCP. O Painel B que exibe os gastos no Ensino Fundamental não mostrou resultados significativos em quase todos os anos (antes e depois da LCP).

A Figura 5 apresenta os resultados para os gastos em educação e para o gasto no ensino fundamental. O gasto em educação aumentou marginalmente em decorrência da Lei da Cota Parte. A estimativa média é um aumento de R\$0,46 centavos por cada real recebido pela cota parte no ano de 2009. No entanto, o gasto no ensino fundamental não apresentou estimativas significativas em quase todos os anos após a introdução da Lei da Cota Parte. Em média o resultado para o ensino fundamental foi um aumento de R\$ 0,23 centavos para cada real distribuído. Assim, a LCP não estimulou de forma semelhante o gasto em educação e não provocou modificações no gasto especificamente do ensino fundamental.

Em suma, os resultados apontam que os municípios que receberam mais recursos com a Lei da Cota Parte gastaram menos do que receberam em educação, seja a educação do ensino fundamental ou não. No entanto, o gasto total aumentou significativamente, mais do que foi inicialmente redistribuído pela cota parte nesses municípios frente aos municípios que não foram beneficiados com a política. Portanto, a Lei da Cota Parte de fato incentivou mais o gasto não educacional do que o gasto em educação para os municípios beneficiados com o programa¹¹.

¹⁰ Existe uma extensa literatura documentando este fenômeno, ver: Hines e Thaler (1995), Inman (2008), Helm e Stuhler (2020). Para o caso do gasto educacional, ver: Gordon (2004) e Cascio et al. (2013). Para uma discussão acerca da literatura nacional ver: Nojosa e Linhares (2018).

¹¹ Foi realizado o mesmo exercício considerando o gasto total subtraído do gasto educacional, definido como gasto não educacional. As estimativas confirmam as conclusões indicando que o gasto não educacional aumentou aproximadamente R\$ 1,45 reais por cada real recebido com a cota parte. Tais resultados não são reportados no artigo por concisão, mas podem ser obtidos por e-mail aos autores.

Importante salientar que este pode ser um resultado esperado pelos formuladores da Lei da Cota Parte pois os recursos distribuídos por tal programa não são vinculados a nenhum tipo de gasto específico, podendo os prefeitos darem o destino que julgarem mais adequado. No entanto, o que as estimativas confirmam é que este tipo de desenho não contribui para aumentar o gasto educacional dos municípios.

4.1.1 Robustez

Nesta subseção são apresentados dois exercícios de robustez para os resultados para os gastos públicos. Primeiro, é adicionado um rico conjunto de variáveis de controle que possam estar associada ao gasto público municipal. Segundo, são substituídos os efeitos fixos associados ao ciclo econômico municipal pelos efeitos fixos municipais variantes no tempo. Adicionalmente, será reportado também o resultado associado a especificação principal usada para estimar as figuras 4 e 5.

A tabela 1 mostra os resultados para os exercícios de robustez. As colunas (1), (4) e (7) são as estimativas considerando a especificação principal. As colunas (2), (5) e (8) mostram os resultados considerando a adição de um conjunto de controles. Esses controles são: PIB municipal, Recursos do Fundo de Participação dos Municípios, Recursos do Fundeb, Razão entre a renda dos 10% mais ricos e dos 40% mais pobres (medida de desigualdade), proporção da população em idade para o ensino fundamental, proporção de pessoas idosas, proporção da população urbana. Todas essas variáveis são indicadas pela literatura como responsáveis por afetar o gasto público em geral e o educacional em particular. Por fim, as colunas (3), (6) e (9) apresentam os resultados ao substituir os efeitos fixos associados ao ciclo econômico por efeitos fixos municipais variantes no tempo.

Não se observa nas estimativas nenhuma diferença significativa de magnitude do impacto sobre os diferentes tipos de gasto. Isso sugere que as estimações principais não são dirigidas por fatores não observados que possam enviesar os resultados. Além disso, não houve diferenças em termos inferenciais, indicando que diferentes especificações não afetam as estimativas dos erros padrões.

Tabela 1: **Resultados de robustez gastos**

Variáveis	Gasto Total			Gasto em Educação			Gasto no Ensino Fundamental		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Transferências Cota Parte	1.627 (0.318)	1.606 (0.303)	1.670 (0.325)	0.381 (0.081)	0.382 (0.082)	0.379 (0.080)	0.258 (0.116)	0.256 (0.119)	0.256 (0.120)
R^2	0.921	0.921	0.911	0.913	0.914	0.911	0.719	0.721	0.713
Observações	2,345	2,291	2,291	2,345	2,291	2,291	2,345	2,291	2,291
Controles Adicionais	N	S	S	N	S	S	N	S	S
Efeito Fixo Ano	S	S	S	S	S	S	S	S	S
Efeito Fixo Municipal	S	S	S	S	S	S	S	S	S
Efeito Fixo Ciclo-por-ano	S	S	N	S	S	N	S	S	N
Efeito Fixo Municipal por Coorte	N	N	S	N	N	S	N	N	S

Erros-Padrão Robustos em parênteses (*** p<0.01, ** p<0.05, * p<0.1).

Outros exercícios de robustez foram realizados, porém não reportados por concisão¹². Os demais exercícios de robustez foram: 1. re-estimação considerando diferentes pontos de corte entre os grupos tratados e controle (30%, 15% e 10% para o grupo tratado de acordo com o ranqueamento do QE_{m2009}) e 2. re-estimação dos erros-padrão considerando procedimento de *bootstrap*. Não houve diferenças relevantes nas conclusões principais.

4.2 Resultados educacionais

4.2.1 Resultados

Como descrito na estratégia empírica será analisado o impacto da interação entre proficiência prévia das escolas e o repasse da cota parte. Contudo o enfoque será nos municípios que foram beneficiados em relação aos municípios que foram prejudicados com a introdução da LCP. Esta diferença será comparada em grupos de escolas com níveis semelhantes de proficiência prévia no 2º ano do Ensino Fundamental. A performance prévia é dividida em quartis e será observado como escolas em cada um desses quartis se diferenciam em 2009 caso o município figure entre os beneficiados ou prejudicados com a LCP. Os resultados estão apresentados em termos de desvios-padrão e são controlados por meio de efeito fixo municipal variante no tempo e efeito fixo temporal.

A tabela 2 apresenta os resultados da estimação da equação (4). Cada coluna compara as escolas em cada uma das categorias de desempenho incorporado a sua classificação na cota parte em relação aos mais beneficiados. O resultado da coluna (1) compara as escolas de baixo desempenho e os municípios mais beneficiados em relação ao mais prejudicados referente a distribuição de cota-parte, o resultado mostra-se não significativo. A coluna (2) compara as escolas de médio desempenho em municípios beneficiados com a cota parte em relação aos prejudicados, o parâmetro de interesse, apesar de maior, continua não significativo. Ambas estimativas implicam que escolas com baixa performance prévia não são afetadas no seu desempenho se elas estão em municípios que se beneficiaram com a LCP.

Já o resultado da coluna (3), relaciona as escolas de médio alto desempenho em municípios beneficiados em relação aos prejudicados em termos de distribuição discricionária de cota parte. Diferentemente dos dois resultados anteriores, a estimativa é positiva, significativa e de elevada magnitude, 1.199 desvios-padrão (p-valor 0.015). Por fim, a comparação entre as escolas de alto desempenho em municípios beneficiados em relação aos prejudicados, é apresentado na coluna (4). A estimativa é, novamente, positiva, significativa e de elevada magnitude 2.144 desvios-padrão (p-valor 0.000). Assim, pode-se concluir que ser beneficiado pela LCP é relevante para as escolas com maior performance educacional prévia.

Tabela 2: **Resultados educacionais**

Variáveis	(1) Q = 1	(2) Q = 2	(3) Q = 3	(4) Q = 4
Tratamento	0.574 (0.473)	0.604 (0.376)	1.199** (0.493)	2.144*** (0.417)
R^2	0.752	0.72	0.663	0.582
Observações	5,730	5,742	6,057	5,472
Efeito Fixo Municipal nor Coorte	S	S	S	S
Efeito Fixo Temporal	S	S	S	S

Erros-Padrão Robustos em parênteses (*** p<0.01, ** p<0.05, * p<0.1)

Nota: O parâmetro tratamento analisa o impacto da interação entre proficiência das escolas e repasse da cota parte previamente a mudança de legislação. Cada Coluna representa um quartil de proficiência (menor desempenho (1) ao melhor desempenho

¹² Tais exercícios podem ser requisitados por e-mail aos autores.

(4), respectivamente). Os quartis são associados ao repasse de LCP (municípios mais beneficiados em relação aos prejudicados) para assim poder definir a variável tratamento.

Em resumo, os sugerem que escolas com baixa performance em 2007 não aumentaram seus resultados educacionais em relação aos municípios que se beneficiaram com a Lei de Cota Parte. Todavia, estar em um município beneficiado com a LCP afeta significativamente o desempenho das melhores escolas. Importante salientar que apesar da Lei de Cota Parte conter mecanismos de controle e penalização para municípios que aumentam a desigualdade entre as escolas ou tentem reduzir a taxa de adesão ao exame, esses mecanismos não estão relacionados a diferenças de performance entre grupos comparáveis de escolas.

4.2.2 Robustez

Nesta subseção são apresentados dois exercícios de robustez para os resultados de desempenho educacional. Uma questão importante associada às estimativas da tabela 2 é a possibilidade de as escolas em diferentes municípios não serem comparáveis entre si em cada um dos quartis. Isto é, escolas com baixa (ou alta) performance em 2007 em municípios que serão beneficiados podem não ser diretamente comparáveis a escolas com baixa (ou alta) performance em municípios que serão prejudicados com a LCP. Para tanto, são realizados dois exercícios de robustez buscando reduzir a possibilidade de incomparabilidade entre as escolas.

O primeiro exercício utiliza uma série de variáveis de controle que possam estar associada ao desempenho futuro das escolas. Tais medidas relacionam-se a qualidade da oferta educacional (como indicadores qualidade dos professores, número de alunos, entre outros) e a características dos estudantes, taxa de alfabetização média, medida de atraso escolar, entre outras. Importante, todas essas covariadas são mensuradas no ano de 2007, antes da introdução da LCP.

O segundo exercício aplica pareamento por entropia (Hainmueller, 2012) considerando as mesmas variáveis pré-determinadas utilizadas no exercício anterior. O pareamento permite atribuir pesos a semelhança entre as escolas, possibilitando uma comparação mais adequada entre tratados e controle. Ambos procedimentos tem o objetivo de permitir que tais escolas sejam comparáveis em cada categoria k .

Tabela 3: **Resultados de robustez educacional**

Painel A: Adição de Controles	(1)	(2)	(3)	(4)
	Q = 1	Q = 2	Q = 3	Q = 4
Tratamento	-1.227 (1.394)	0.582 (0.487)	-4.705 (6.298)	1.924*** (0.391)
R^2	0.777	0.745	0.703	0.646
Observações	4,496	4,683	4,939	4,531
Painel B: Pareamento por Entropia	(1)	(2)	(3)	(4)
	Q = 1	Q = 2	Q = 3	Q = 4
Tratamento	0.26 (0.822)	0.557 (0.373)	1.236** (0.583)	2.128*** (0.383)
R^2	0.765	0.731	0.69	0.612
Observações	4,496	4,683	4,939	4,531
Efeito Fixo Municipal nor Coorte	S	S	S	S
Efeito Fixo Temporal	S	S	S	S

Erros-Padrão Robustos em parênteses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

Nota: O parâmetro tratamento analisa o impacto da interação entre proficiência das escolas e repasse da cota parte previamente a mudança de legislação. Cada Coluna representa um quartil de proficiência (menor desempenho (1) ao melhor desempenho (4), respectivamente). Os quartis são associados ao repasse de LCP (municípios mais beneficiados em relação aos prejudicados) para assim poder definir a variável tratamento.

Os resultados apresentados na tabela 3, disponível acima, estão divididos em dois painéis. O painel A apresenta a especificação com a adição de variáveis de controle¹³. O resultado para esse primeiro teste de robustez se mostra não significativo, com exceção da categoria 4, apresentado na coluna (4), em que as melhores escolas em termos de desempenho e municípios mais beneficiados apresentam resultado positivo de magnitude 1.924 desvios-padrão (p-valor 0.000), semelhante ao encontrado na primeira estimação.

O resultado do painel B refere-se a aplicação do pareamento por entropia. Assim como aos resultados principais, tais estimativas são positivas nas colunas (1) e (2), porém não significativas. Da mesma forma, os resultados para as escolas com melhores performances são positivos e significativos, semelhante as conclusões principais. A magnitude dos resultados é marginalmente menor, contudo bastante expressiva, acima de 2 desvios padrões para coluna (4).

Os resultados sugerem que as estimações não são potencialmente digeridas por fatores observados que podem enviesar os resultados. Como forma de explorar e dar uma maior evidência aos resultados já encontrados, é feito mais um exercício, dessa vez, mudando a variável de resultado, para variáveis relacionadas a maiores gastos em educação

4.2.3 Mecanismos

Uma possível explicação para os resultados anteriores é a hipótese de que os prefeitos estão alocando seus recursos de forma diferenciada nas escolas com diferentes categorias. Aquelas escolas que tem melhor produtividade com o uso dos recursos públicos podem estar sendo preferida pelos gestores em detrimento das escolas com menor produtividade. Como forma de verificar esta hipótese de alocação diferenciada de recursos, estimou-se um modelo tal como a equação (2) tendo como variáveis de resultado a quantidade de Horas Aula Média por Dia e a Média de Alunos por Turma. Ambas as variáveis estão associadas a oferta e a qualidade da educação pública, bem como maiores gastos em educação.

Os resultados estão divididos em dois painéis e são apresentados na tabela 4. O Painel A refere-se ao modelo em que a variável de resultado é horas diárias de aulas. Utiliza-se a mesma técnica de pareamento por entropia para evitar a incomparabilidade entre as escolas. Os resultados mostram que não há diferenças significativas entre as escolas, porém, as melhores escolas tem estimativa pontual de 1.5 horas diárias adicionais que as escolas em municípios prejudicados com a LCP.

Já no Painel B, a variável de resultado é o tamanho médio das turmas, também controlando pelo pareamento por entropia. As estimativas são não significativas para as especificações em (1), (2), (3). Entretanto, a estimativas para as melhores escolas indica que tais escolas possuem em média 7.6 alunos a menos que escolas comparáveis em municípios prejudicados com a LCP. Estes resultados complementares aos anteriores sugerem que houve uma alocação diferenciada dos recursos disponíveis nos municípios em escolas com melhor desempenho prévio.

¹³ Variáveis de controle: % de alunos com alfabetização incompleta, % de alunos com alfabetização intermediária, % de alunos com alfabetização suficiente, % de alunos com alfabetização desejável, media dos professores em determinada escola com especialização, média de professores em determinada escola com pós graduação (mestrado e/ou doutorado), média dos docentes em determinada escola dividida em 4 categorias (até 24 anos, 24 a 29 anos, 30 a 39 anos, 40 a 49 anos), sexo, cor/raça branco, aprovação e abandono no 1º e 2º ano, gasto total per capita, PIB per capita e fundo de participação municipal per capita. Todas as variáveis para 2008.

Ou seja, escolas que já apresentam elevada performance anterior a LCP tiveram uma redução no número de alunos por turma quando comparadas a escolas de performance semelhante em municípios prejudicados com a LCP. Não houve diferença entre as escolas de baixa performance antes da LCP.

Tabela 4 : **Resultados Mecanismos**

Painel A: Horas Diárias de Aula	(1) Q=1	(2) Q=2	(3) Q=3	(4) Q=4
Tratamento	-	-0.248	-0.02	1.501
	-	(0.000)	(0.013)	(1.508)
R^2	-	0.708	0.6	0.628
Observações	-	3,370	3,824	3,612
Painel B: Tamanho Médio das Turmas	(1) Q=1	(2) Q=2	(3) Q=3	(4) Q=4
Tratamento	12.104	3.399	0.795	-7.625**
	(7.585)	(0.000)	(5.071)	(3.041)
R^2	0.523	0.487	0.396	0.467
Observações	2,920	3,347	3,797	3,589
Balanceamento por Entropia	S	S	S	S
Efeito Fixo Municipal por Coorte	S	S	S	S
Efeito Fixo Temporal	S	S	S	S

Erros Padrão Robustos em parênteses (*** p<0.01, ** p<0.05, * p<0.1)

5 CONCLUSÕES

Este artigo analisou o impacto da introdução da Lei da Cota Parte (LCP) em dois diferentes aspectos: alocação das despesas públicas municipais e desigualdade de desempenho no 2º ano do ensino fundamental. Os resultados mostram que os municípios beneficiados com LCP destinaram maiores recursos para gastos não educacionais. Os gastos educacionais totais aumentaram menos que o valor recebido com a cota parte e, especificamente, o gasto no ensino fundamental não foi afetado. Tal resultado sugere que a LCP não induziu os municípios a aumentarem seus gastos em educação.

Posteriormente foi verificado se o impacto educacional da LCP foi diferenciado entre as escolas com diferentes performances prévias a introdução do programa. Os resultados mostram que apenas as escolas com melhores desempenhos prévios melhoram seus resultados quando estavam em municípios beneficiados pela LCP. Isso sugere a existência de uma alocação diferenciada de recursos nos municípios beneficiados. Em análise de mecanismos foi verificado a possibilidade dessa alocação diferenciada de recursos.

Tomados em conjunto, os resultados apontam que o efeito da LCP não induziu os municípios a um gasto maior em educação e aumentou a desigualdade educacional a nível da escola. Quanto ao primeiro ponto, percebe-se que os municípios quase dobraram seus gastos totais frente a um aumento de recursos da LCP, efeito este conhecido como Flypaper. Além disso, tal elevação no gasto total é persistente no tempo, sugerindo que a LCP produziu modificações de longo prazo na quantidade de recursos totais despendidos pelo município.

Quanto ao segundo ponto, os resultados mostram que as escolas com baixa performance em 2007 não aumentaram seus resultados educacionais em relação aos municípios que se beneficiaram com a LCP. A principal explicação é a presença de uma alocação de recursos intramunicipalmente direcionada para as escolas que têm maior potencial de gerar bons resultados nos testes padronizados.

Alguns exercícios foram realizados, afim de dar maior robustez aos resultados, um para o gasto e dois educação. Para os resultados do Flypapaer é adicionado um conjunto de variáveis de controle que possam estar associada ao gasto público municipal. Bem como, são substituídos os efeito fixos associados ao ciclo econômico municipal pelos efeitos fixos municipais variantes no tempo. Não observando diferença significativa de magnitude do impacto sobre os diferentes tipos de gasto.

Para os resultados educacionais utiliza-se uma série de variáveis de controle que possam estar associada ao desempenho futuro das escolas. aplica pareamento por entropia (Hainmueller, 2012) considerando as mesmas variáveis pré-determinadas utilizadas para o controle anterior. Os resultados sugerem que as estimações não são potencialmente digeridas por fatores observados que podem enviesar os resultados. A última atividade, mudando a variável de resultado, para variáveis relacionadas a maiores gastos em educação, sugerindo mais uma vez não haver diferença entre as escolas de baixa performance antes da LCP

Os resultados encontrados são intrigantes e nos leva a investigar quais outros canais foram chaves para melhorar o desempenho educacional cearense, bem como, qual a significância da LCP no conjunto de mecanismos educacionais existentes no Ceará. Tais questões não são alvo desta pesquisa, sendo deixadas para pesquisas futuras.

REFERÊNCIAS

BRANDÃO, Júlia Barbosa. **O rateio de ICMS por desempenho de municípios no Ceará e seu impacto em indicadores do sistema de avaliação da educação**. 2014. Tese de Doutorado.

BRASIL. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. Diário Oficial da União, Brasília, DF, 23 dez. 1996. Seção 1, p. 27833.

CARNEIRO, Diego; IRFFI, Guilherme. PROBLEMA DO RISCO MORAL NA EDUCAÇÃO BÁSICA: UM MODELO DE AGENTE-PRINCIPAL PARA A DISTRIBUIÇÃO DE RECURSOS DA COTA PARTE DO ICMS. In: **Anais do XLIV Encontro Nacional de Economia [Proceedings of the 44th Brazilian Economics Meeting]**. ANPEC-Associação o Nacional dos Centros de Pós-Graduação em Economia [Brazilian Association of Graduate Programs in Economics], 2018.

CASCIO, Elizabeth U.; GORDON, Nora; REBER, Sarah. Local responses to federal grants: Evidence from the introduction of Title I in the South. **American Economic Journal: Economic Policy**, v. 5, n. 3, p. 126-59, 2013. CEARÁ. Decreto nº 30.796, de 29 de dezembro de 2011. Altera o dispositivo do Decreto nº 29.881, de 31 de agosto de 2009, e dá outras providências.

_____. Lei nº 12.612, de 7 de agosto de 1996. Define, na forma do Art. 158, Parágrafo Único, II, da Constituição Federal, critérios para distribuição da parcela de receita do produto de arrecadação do ICMS pertencente aos municípios.

_____. Lei nº 14.023, de 17 dezembro de 2007. Modifica dispositivos da Lei nº 12.612, de 7 de agosto de 1996, que define critérios para distribuição da parcela de receita do produto e

arrecadação do Imposto Sobre Operações Relativas à Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação – ICMS, pertencente aos municípios e dá outras providências.

CILLIERS, Jacobus; MBITI, Isaac M.; ZEITLIN, Andrew. Can public rankings improve school performance? Evidence from a nationwide reform in Tanzania. **Journal of Human Resources**, p. 0119-9969R1, 2020.

CULLEN, Julie Berry; REBACK, Randall. **Tinkering toward accolades: School gaming under a performance accountability system**. Emerald Group Publishing Limited, 2006.

FERNANDES, Mauricio M.; FERRAZ, Claudio. **Conhecimento ou Práticas Pedagógicas? Medindo os Efeitos da Qualidade dos Professores no Desempenho dos Alunos**. Texto para discussão, 2014.

FIGLIO, David N.; GETZLER, Lawrence S. Accountability, ability and disability: Gaming the system. **Advances in applied microeconomics**, v. 14, p. 35-49, 2006.

FIGLIO, David N.; LUCAS, Maurice E. What's in a grade? School report cards and the housing market. **American economic review**, v. 94, n. 3, p. 591-604, 2004.

FIGLIO, David N.; ROUSE, Cecilia Elena. Do accountability and voucher threats improve low-performing schools?. **Journal of Public Economics**, v. 90, n. 1-2, p. 239-255, 2006.

GORDON, Nora. Do federal grants boost school spending? Evidence from Title I. **Journal of Public Economics**, v. 88, n. 9-10, p. 1771-1792, 2004.

HADDAD, Mônica A.; FREGUGLIA, Ricardo; GOMES, Cláudia. Public spending and quality of education in Brazil. **The Journal of Development Studies**, v. 53, n. 10, p. 1679-1696, 2017.

HAINMUELLER, Jens. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. **Political analysis**, p. 25-46, 2012.

HANUSHEK, Eric A. Why quality matters in education. **Finance and development**, v. 42, n. 2, p. 15-19, 2005.

HELM, Ines; STUHLER. **The Dynamic Response of Municipal Budgets to Revenue Shocks**. Working paper.2020.

HINES, James R.; THALER, Richard H. The flypaper effect. **Journal of economic perspectives**, v. 9, n. 4, p. 217-226, 1995.

JACOB, Brian A. Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. **Journal of public Economics**, v. 89, n. 5-6, p. 761-796, 2005.

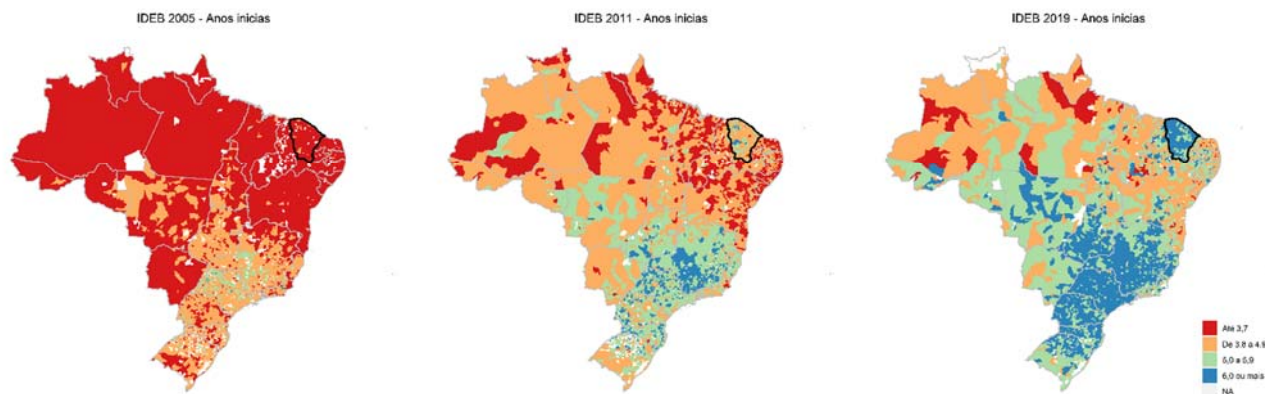
JACKSON, C. Kirabo; JOHNSON, Rucker C.; PERSICO, Claudia. The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. **The Quarterly Journal of Economics**, v. 131, n. 1, p. 157-218, 2016.

JACKSON, C. Kirabo; WIGGER, Cora; XIONG, Heyu (2020a). Do school spending cuts matter? Evidence from the great recession. **American Economic Journal: Economic Policy**. forthcoming.

- INMAN, Robert P. The flypaper effect. **National Bureau of Economic Research**, 2008.
- LEE, Jessica D.; MEDINA, Octavio. **Results-based financing in education: Learning from what works**. World Bank, 2019.
- LOUREIRO, Andre et al. The State of Ceara in Brazil is a Role Model for Reducing Learning Poverty. 2020.
- MANUELLI, Rodolfo E.; SESHADRI, Ananth. Human capital and the wealth of nations. **American economic review**, v. 104, n. 9, p. 2736-62, 2014.
- MBITI, Isaac et al. Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania. **The Quarterly Journal of Economics**, v. 134, n. 3, p. 1627-1673, 2019.
- MURALIDHARAN, Karthik; SINGH, Abhijeet; GANIMIAN, Alejandro J. Disrupting education? Experimental evidence on technology-aided instruction in India. **American Economic Review**, v. 109, n. 4, p. 1426-60, 2019
- NEAL, Derek; SCHANZENBACH, Diane Whitmore. Left behind by design: Proficiency counts and test-based accountability. **The Review of Economics and Statistics**, v. 92, n. 2, p. 263-283, 2010.
- NOJOSA, Glauber Marques et al. Variabilidade do efeito flypaper e força política: uma análise para os municípios brasileiros. In: **Anais do XLIV Encontro Nacional de Economia [Proceedings of the 44th Brazilian Economics Meeting]**. ANPEC-Associação Nacional dos Centros de Pós-Graduação em Economia [Brazilian Association of Graduate Programs in Economics], 2018.
- PETTERINI, Francis Carlo; IRFFI, Guilherme Diniz. Evaluating the impact of a change in the ICMS tax law in the state of Ceará in municipal education and health indicators. **Economia**, v. 14, n. 3-4, p. 171-184, 2013.
- REBACK, Randall. Teaching to the rating: School accountability and the distribution of student achievement. **Journal of public economics**, v. 92, n. 5-6, p. 1394-1415, 2008.
- SHIRASU, Maitê Rimekká; IRFFI, Guilherme Diniz; PETTERINI, Francis Carlo. Melhorando a qualidade da educação por meio do incentivo orçamentário aos prefeitos: o caso da Lei do ICMS no Ceará.
- SIMÕES, Armando Amorim; ARAÚJO, Erika Amorim. O ICMS e sua potencialidade como instrumento de política educacional. **Cadernos de Estudos e Pesquisas em Políticas Educacionais**, v. 3, p. 48-48, 2019.
- ZIDAR, Owen. Tax cuts for whom? Heterogeneous effects of income tax changes on growth and employment. **Journal of Political Economy**, v. 127, n. 3, p. 1437-1472, 2019.

APÊNDICE

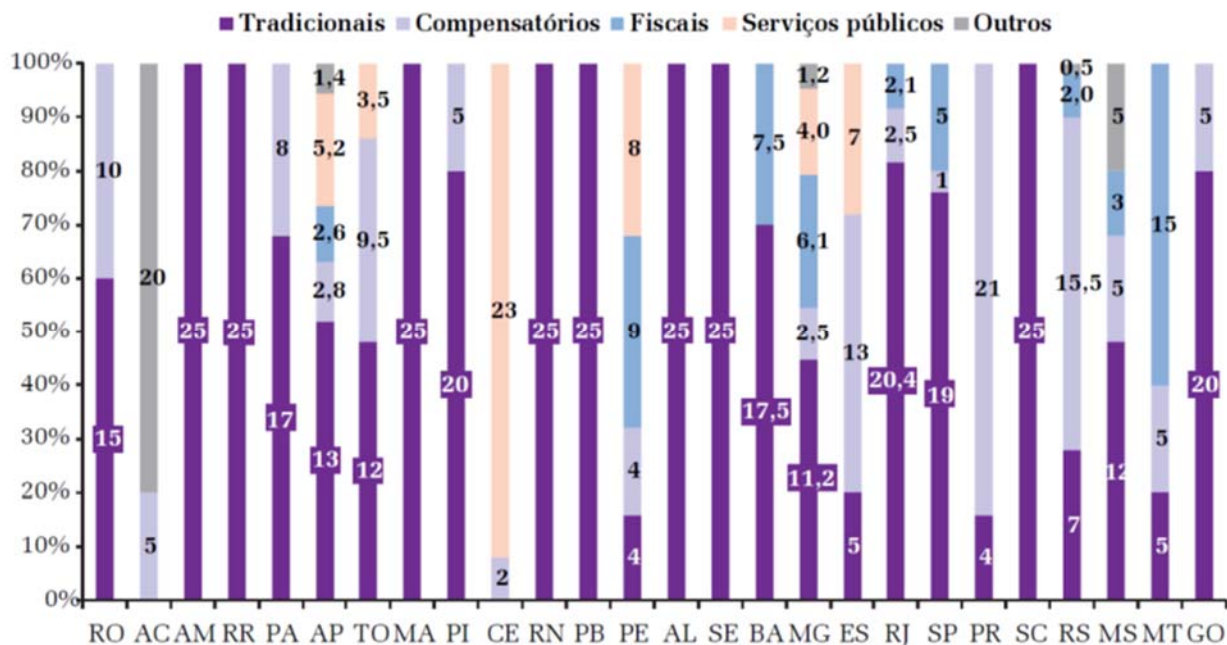
Figura 1: Evolução média dos municípios para IDEB nos anos iniciais



Nota: O conjunto de mapas mostra a evolução do IDEB para os anos iniciais ao longo dos anos de 2005, 2011, 2019. O estado do Ceará registrou a maior evolução histórica, passando de 2,8 em 2005 para 6,3 em 2019. Além disso, 131 municípios cearenses atingiram média 6 para essa etapa de ensino no ano de 2019. Ainda conforme o indicador, no ano de 2019, 21 municípios e 79 escolas estão entre os 100 mais bem avaliados nas séries iniciais.

Fonte: Elaboração Própria, com os microdados do SAEB.

Figura 2: Distribuição da parte Discrecionária de ICMS pelos estados



Nota: A figura 1 apresenta as formas adotadas pelos estados para distribuir recursos da cota parte do ICMS. **Tradicionais** referentes ao VAF, parte igualitária, população e área geográfica; **Compensatórios** que visam ressarcir determinados municípios de situações ou atividades que não integram a base de incidência do ICMS; **Fiscais** que levam em conta a receita tributária própria municipal e algumas medida de carência de recursos com vistas à equalização da capacidade orçamentária municipal; **Serviços públicos** incluem as políticas de educação, saúde e saneamento; **Outros** contemplam entre os programas estaduais de caráter bem particular como a preservação do patrimônio cultural, atividades ligadas ao esporte e ao turismo, bem como o número de eleitores.

Fonte: SIMÕES; ARAÚJO (2019).

Quadro 1: Descrição das Variáveis Utilizadas para Resultados Educacionais

Variável	Descrição
Taxa de Aprovação	Proporção de alunos de 1º e 2º Ano de Ensino Fundamental aprovados.
Taxa de Reprovação	Proporção de alunos de 1º e 2º Ano de Ensino Fundamental reprovados
Taxa de Abandono	Proporção de alunos de 1º e 2º Ano de Ensino Fundamental que abandonaram a escola
Distorção Idade-Série	Proporção de alunos de 1º e 2º Ano do Ensino Fundamental com idade acima da adequada a série que frequenta
Nota SPAECE-Alfa	Proficiência média dos alunos do 2º Ano do Ensino Fundamental no SPAECE - Alfa
Não Alfabetizados	Porcentagem de alunos não alfabetizados
Alfabetização Incompleta	Porcentagem de alunos com alfabetização Incompleta
Intermediário	Porcentagem de alunos com alfabetização Intermediária
Suficiente	Porcentagem de alunos com alfabetização Suficiente
Desejável	Porcentagem de alunos com alfabetização Desejável
Docentes com Ensino Superior	Média dos docentes com escolaridade de nível superior
Branco	Média dos docentes declarados com cor/raça Branco
Sexo	Média dos docentes do sexo Feminino
Idade1	Média de Idade dos docentes até 24 anos
Idade2	Média de Idade dos docentes de 24 a 29 anos
Idade3	Média de Idade dos docentes de 30 a 39 anos
Idade4	Média de Idade dos docentes de 40 a 49 anos
Idade5	Média de Idade dos docentes acima de 50 anos
duração aula	Média de horas aula da escola no 2º ano
especialização	Média dos docentes com especialização
pós	Média dos docentes com mestrado e/ou doutorado
sem pós	Média dos docentes sem nenhuma pós graduação

QUEM SÃO OS JOVENS ‘NEM NEM’ NA REGIÃO METROPOLITANA DE FORTALEZA

WHO ARE YOUNG PEOPLE ‘NEET’ IN THE METROPOLITAN REGION OF FORTALEZA

Antonia Jaine da Silva Pereira*

Silvana Nunes de Queiroz**

RESUMO

Este estudo tem como objetivo traçar o perfil sociodemográfico dos jovens nem nem que residem na Região Metropolitana de Fortaleza (RMF), além de estimar a probabilidade de estarem nessa condição. Os microdados da PNAD 2015 são a principal fonte de informações. Quanto aos resultados, as características mais marcantes no perfil do jovem nem nem de 15 a 29 anos, é a predominância de mulheres, da raça/cor parda, com ensino médio completo ou superior incompleto, solteiro, mora com os pais, em área urbana e em domicílios com rendimento até ½ salário mínimo. O modelo probit mostrou que somente a escolaridade afeta negativamente a probabilidade de estar nessa condição, enquanto as demais variáveis tem efeito ‘positivo’, ao aumentar a chance de ser nem nem, com destaque para o rendimento familiar per capita abaixo de meio salário mínimo, variável explicativa com maior impacto, reafirmando a vulnerabilidade financeira dos jovens que estão em tal condição.

Palavras-chave: Educação, Trabalho, Jovem, Nem Nem, Região Metropolitana de Fortaleza

ABSTRACT

This study aims to outline the socio-demographic profile of young people neet, even those who live in the Metropolitan Region of Fortaleza (RMF), in addition to estimating the probability of being in this condition. PNAD 2015 microdata is the main source of information. As for the results, the most striking characteristics in the profile of the young people neet, from 15 to 29 years old, is the predominance of women, of mixed race/color, with complete high school or incomplete higher education, single, lives with their parents, in an urban area and in households with income up to ½ minimum wage. The probit model showed that only schooling negatively affects the probability of being in this condition, while the other variables have a 'positive' effect, by increasing the chance of being neet, with emphasis on per capita family income below half the minimum wage, explanatory variable with the greatest impact, reaffirming the financial vulnerability of young people who are in such condition.

Keywords: Education, Job, Young, Neet, Metropolitan Region of Fortaleza

Área 2 - Economia Social.

Classificação JEL: J38; B55; J13.

* Economista (URCA) e mestranda do Programa de Pós-Graduação em Demografia (PPGDEM/UFRN).

E-mail: jainerodrigues22@yahoo.com. Tel.: (84) 99835 6231.

** Professora do Departamento de Economia da Universidade Regional do Cariri (URCA) e do Programa de Pós-Graduação em Demografia (PPGDEM/UFRN). E-mail: silvanaqueirozce@yahoo.com.br. Tel.: (88) 9 9792 8808.

1 INTRODUÇÃO

Assistiu-se, nos últimos anos, significativo crescimento no contingente absoluto e relativo de jovens de 15 a 29 anos que não estudam e não trabalham. Tal tendência tem preocupado os formuladores de políticas públicas e despertado o interesse de pesquisadores, já que não é um fenômeno recente e nem se restringe ao Brasil. Com isso, o tema passou a ganhar atenção entre os estudiosos, tornando-se pauta de pesquisa a partir dos anos 2000, com os quais tais jovens passaram a ser denominados “nem nem” (nem estuda, nem trabalha).

Trata-se de um tema que se tornou foco de discussões, contudo, voltados notadamente para o jovem nem nem no Brasil (CAMARANO et al., 2006; FUNDAÇÃO PERSEU ABRAMO, 2013; COSTA; ULYSSEA, 2014). Assim, apesar da importância da temática, não se conhece até o presente momento estudos específicos sobre esses jovens no estado do Ceará.

Diante do exposto, este estudo pretende preencher esta lacuna, ao traçar o perfil sociodemográfico dos jovens nem nem residentes na Região Metropolitana de Fortaleza (RMF). Além disso, é feita uma descrição dos investimentos públicos realizados em educação e programas para o primeiro emprego no Ceará, a partir dos anos 2000.

Afora a contribuição científica para a discussão do tema, este estudo pode interessar aos formuladores de políticas públicas, ao mostrar quem são os jovens nem nem residentes na RMF, e nortear a atuação do governo em pontos como o combate à evasão escolar e ao desemprego, ou ainda o auxílio para a saída da condição de nem nem.

Para alcançar o objetivo proposto, utilizou-se como principal fonte de dados os microdados da Pesquisa Nacional por Amostra de Domicílio (PNAD), disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A extração dos dados procedeu com o software estatístico SPSS (Statistical Package for the Social Science) que permitiu a obtenção das informações para as estatísticas descritivas, e através do software STATA 13, foi possível a construção do modelo probit. Também fez-se uso do programa Excel para a construção de tabelas e o QGis (versão 2.18.12) para a elaboração do mapa de localização.

No tocante a organização do trabalho, além desta introdução, o mesmo conta com mais cinco seções. A segunda faz uma breve descrição dos investimentos públicos realizados em educação a partir dos anos 2000, considerando ações realizadas no Ceará. Já a terceira seção descreve os programas para o primeiro emprego efetuados pelo governo cearense. Por sua vez, a quarta seção apresenta os procedimentos metodológicos para o alcance dos objetivos propostos. Na quinta seção são apresentados e analisados os dados acerca do perfil sociodemográfico do jovem nem nem na RMF, enquanto a sexta seção destaca as considerações finais do estudo.

2. INVESTIMENTOS EM EDUCAÇÃO NO CEARÁ

O Ceará, estado onde localiza-se a RMF, tem buscado melhorias na educação através de diversos programas e investimentos, tornando-se um tema sempre presente na agenda do Governo (COELHO, 2013). No início dos anos 2000, Naspolini (2001) no estudo sobre a reforma da educação básica no Ceará afirma:

[...] o acesso não é mais o problema fundamental, pois 98% das crianças e adolescentes de 7 a 14 anos estão matriculadas nas escolas. Mas não é apenas com as tradicionais providências relacionadas com o acesso que será resolvido o problema das crianças que continuam fora da escola. O caso é mais complexo. É de inclusão, o que implica em ações mais abrangentes de política social. Simultaneamente, são necessárias providências para garantir a permanência e o sucesso escolar, relacionados, portanto, com a qualidade. Nesse sentido, o Estado organizou e

implantou o ensino em Ciclos de Formação, criou as Classes de Aceleração e vem desenvolvendo um Sistema de Acompanhamento Pedagógico envolvendo toda a rede pública de ensino (NASPOLINI, 2001, p. 174).

A partir de meados da mesma década foram implementados inúmeros projetos, programas, prêmios e ações, envolvendo ensino, transporte (Transporte Escolar), saúde (Programa Saúde na Escola - PSE), alimentação (Programa Nacional de Alimentação Escolar - PNAE/CE) e esporte (Programa de Formação Esportiva Escolar - Atleta na Escola), além de programas de conscientização e proteção (Programa Escola que Protege - PETECA, Programa Geração da Paz). Há ainda programas e prêmios de estímulo aos gestores, professores e estudantes (Prêmio Escola Nota 10, Professor Aprendiz, Projeto Jovem de Futuro), e aqueles que incentivam a continuação dos estudos (Projeto Rumo à Universidade, Enem Chego Junto Chego Bem, Eu Curto a Universidade) e a formação profissional (Escola Estadual de Ensino Profissional - EEEP) (CEARÁ, 2016).

Entre os programas voltados para o ensino destaca-se o PAIC (Programa de Alfabetização na Idade Certa) que, segundo Coelho (2013, p. 4), é um “programa prioritário na agenda estratégica do Governo do Estado do Ceará, instituído no ano de 2007, para fazer frente ao desafio de eliminar o problema do analfabetismo escolar”.

Inicialmente desenvolvido para alguns municípios cearenses e assumido em 2007 como política pública do Ceará, o PAIC figura como importante ferramenta para a alfabetização das crianças da rede pública de ensino. Em 2011, o Governo do Estado lança o PAIC MAIS, estendendo as ações do programa até o 5º ano do ensino fundamental em todas as escolas públicas cearenses. Atualmente o PAIC conta com a parceria de nove instituições: Associação para o Desenvolvimento dos Municípios do Estado do Ceará - APDMCE, Associação dos Municípios do Estado do Ceará - APRECE, Fórum de Educação Infantil do Ceará, Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico - FUNCAP, Secretaria da Cultura do Estado do Ceará - SECULT, Secretaria da Educação do Estado do Ceará - SEDUC, União dos Dirigentes - UNDIME, UNICEF (Fundo das Nações Unidas para a Infância) e União Nacional dos Conselhos Municipais de Educação - UNCME (PAIC, 2020).

Em um estudo sobre a avaliação do ensino através do SPAECE (Sistema Permanente de Avaliação da Educação Básica do Ceará) para os anos de 2007 a 2010, Lima afirma que:

Os resultados produzidos nas avaliações desses quatro anos demonstram uma tendência acentuada e contínua de melhoria do desempenho em todas as séries e disciplinas avaliadas. Verifica-se que a melhoria no desempenho escolar é mais acentuada nas primeiras séries do ensino fundamental, particularmente no 2º ano. Já no 5º ano, embora tenha apresentado sinais de melhoria, constata-se ainda a existência de defasagens em relação aos níveis desejados para essa etapa de escolaridade. Por outro lado, esses resultados vislumbram, a médio e longo prazo, a possibilidade de melhoria dos desempenhos nos anos subsequentes, uma vez que se espera que os alunos tenham consolidado as habilidades básicas de leitura e escrita (LIMA, 2012, p. 57).

Há também no campo das políticas públicas afinidade entre a Educação e a Saúde, já que se baseiam na universalização de direitos fundamentais (BRASIL, 2009). Dentro dessa perspectiva tem-se o Programa Saúde na Escola (PSE), cuja implantação em uma escola do município de Fortaleza tornou possível aos estudantes a obtenção de hábitos saudáveis a partir das informações recebidas, além de trazer para os profissionais da saúde a percepção acerca do seu papel social de educador (SANTIAGO et al., 2012).

Ainda nesse campo, podem ser citados o Programa Nacional de Alimentação Escolar (PNAE/CE) e o Programa de Formação Esportiva Escolar (Atleta na Escola). Sobre o primeiro, este tem como principal objetivo oferecer aos alunos uma alimentação saudável, contribuindo

para o crescimento, desenvolvimento e rendimento, além de incentivar a prática de hábitos alimentares saudáveis (CEARÁ, 2016). Além disso, avaliar e acompanhar o estado nutricional de alunos, professores e demais membros da comunidade escolar é outra das ações previstas no PSE (BRASIL, 2009). Quanto a inserção em atividades esportivas no âmbito escolar, o primeiro objetivo era a busca pela democratização do acesso, mas também o incentivo a prática de esportes ou ainda revelar talentos que poderiam ser orientados pela escola (CEARÁ, 2016).

Em se tratando dos programas e prêmios de estímulo aos gestores, professores e estudantes, são exemplos o Prêmio Escola Nota Dez, o Professor Aprendiz e o Projeto Jovem de Futuro, respectivamente. O Prêmio Escola Nota Dez foi criado inicialmente em 2009, mas vige atualmente sob a Lei Nº 15.923 de 15 de dezembro de 2015, tendo como objetivo a valorização da gestão educacional voltada para a aprendizagem do aluno. As premiações baseiam-se nos Índices de Desempenho Escolar (IDE) dos alunos do 2º, 5º e 9º ano, obtidos através das avaliações do SPAECE. O valor da premiação é calculado para cada escola, tanto as de maior quanto as de menor resultado, e estas recebem apenas uma parte (75% e 50%, respectivamente), sendo o recebimento do restante condicionado ao apoio entre os dois tipos, para que seja mantido ou elevado o resultado daquela que se saiu melhor, e atingida determinada meta no caso daquelas que não se saíram bem. O intuito é estimular a cooperação técnica e pedagógica entre estas, dando as escolas premiadas a oportunidade de desenvolver ações em apoio às demais, possibilitando o desenvolvimento e manutenção do trabalho de qualidade educacional no âmbito do sistema público de ensino no Estado (CEARÁ, 2016).

Já o Professor Aprendiz trata-se da criação de um ambiente favorável ao desenvolvimento de pesquisas por parte dos professores e estudantes, possibilitando ainda a apresentação/publicação das mesmas. Assim, decidiu-se pela iniciativa de realizar colóquios. O evento teve cinco edições, sendo a última em 2013, com a temática “Diversidade, Qualidade na Educação e Acesso à Educação Superior” (CEARÁ, 2016).

O Projeto Jovem de Futuro, por sua vez, é desenvolvido através de uma parceria entre a Secretaria de Educação do Ceará e o Instituto Unibanco, que oferece apoio técnico e financeiro às escolas participantes para o melhoramento do seu desempenho. As escolas participantes são de ensino médio e recebem, durante um período de 3 anos, capacitação e assessoria técnica, além de financiamento para a execução de planos estratégicos para o melhoramento. Desse modo, o projeto figura como uma proposta à mobilização de todos, sejam alunos, professores ou famílias, em torno de metas pactuadas para atender um objetivo comum, que é a garantia da entrada, permanência com bom desempenho e término do ensino médio.

Além desses, ainda há os programas que incentivam a continuação dos estudos, como o projeto Rumo à Universidade e o Enem Chego Junto Chego Bem, ou ainda a formação profissional, como ocorre nas Escolas Estaduais de Ensino Profissional (EEEP). O Projeto Rumo à Universidade é basicamente de preparação, buscando ampliar as oportunidades de acesso às universidades públicas através da disponibilização de materiais didáticos, do auxílio em programas de estudo etc. O Enem Chego Junto Chego Bem parte também da premissa de suporte para o ingresso no ensino superior, especificamente para a realização do Exame Nacional do Ensino Médio (ENEM), mas foi se desenvolvendo e passou a abranger também a organização de transporte, hospedagem e alimentação, além das atividades pedagógicas (CEARÁ, 2016).

Já as Escolas Estaduais de Ensino Profissional (EEEP) começaram a ser instituídas no Ceará a partir de 2008 e somam no ano de 2016, 116 em todo o estado, com o objetivo de tornar possível um futuro mais igual e com mais oportunidades para os jovens cearenses. Para isso, as Escolas Estaduais funcionam em tempo integral e integram ao ensino médio cursos técnicos profissionalizantes, ampliando as perspectivas do jovem cearense ao articular os direitos à educação e ao trabalho. Os recursos utilizados para a implantação da rede de ensino profissional

provêm do Governo do Estado do Ceará e do apoio do Governo Federal por meio do Programa Brasil Profissionalizado (CEARÁ, 2016).

Acompanhando o desenvolvimento dos diversos projetos e programas, é necessário saber sobre os métodos de avaliação do ensino. Nesse quesito destaca-se o Sistema Permanente de Avaliação da Educação Básica do Ceará (SPAECE), que consiste em uma avaliação censitária e universal, que abrange escolas estaduais e municipais, e é composta por itens de Língua Portuguesa e Matemática, elaborados pelos professores da rede pública. O sistema vem sendo implementado desde 1992, mas a partir de 2007 foi modificado, ampliando a avaliação do ensino médio para os três anos. Assim, tem-se avaliações desde a alfabetização, com o SPAECE - Alfa realizado no 2º ano, passando pelo ensino fundamental com avaliações no 5º e no 9º ano, e continuando no ensino médio, que passou a ser avaliado em seus três anos a partir de 2007 (CEARÁ, 2020b).

Quanto aos resultados obtidos com as ações do governo na área educacional, estes foram positivos em relação à qualidade do ensino, que segundo Naspolini (2001) tornou-se o principal problema após o esforço de universalização do acesso à educação. Segundo dados do INEP, o IDEB cearense referente ao Fundamental I chegou em 2013 a 5,2, e em 2015 atingiu 5,9. Com esse resultado o estado já ultrapassou sua meta para 2021 (5,4) e fica a um décimo de atingir a do país (6,0), para o mesmo ano. O Fundamental II, avaliado na 8ª série/9º ano, também vai bem segundo o IDEB, pois em 2015 (4,8) já havia ultrapassado a meta para 2017 (4,3). Já as avaliações do 3º ano do ensino médio, mostram avanço lento, pois até 2011 sua meta havia sido atingida (3,6), mas a partir desse ano apresentou avanço lento, distanciando-se das metas dos anos seguintes (INEP, 2020). Entretanto, como salienta Lima (2012), há boas perspectivas para as séries seguintes, supondo a consolidação do aprendizado nos anos iniciais, como aponta o IDEB.

Além do ensino médio, a educação superior é outra fase em que se encontram boa parte dos jovens ‘nem nem’ aqui estudados, tendo interrompido ou já concluído a graduação. Com relação a este nível de ensino, podem ser elencadas várias ações quanto ao seu surgimento e distribuição pelo estado.

De início cabe destacar que a criação de cursos de graduação no Ceará se iniciou tardiamente em relação a outros estados, e prosseguiu de forma lenta, dado que sua primeira universidade (Universidade Federal do Ceará - UFC) surge já na segunda metade do século XX, porquanto em outras UF's são criadas no século XIX. Quanto à iniciativa privada, “em meio à expansão universitária ocorrida após a reforma universitária de 1968, apenas duas instituições particulares foram criadas nesse período na capital cearense, ambas em 1971: a Faculdade de Filosofia de Fortaleza e a Universidade de Fortaleza (UNIFOR)” (SULIANO, 2013, p. 43).

Já as primeiras indicações de interiorização do ensino superior no estado foram observadas em 1960, quando foi criada a Faculdade de Filosofia do Crato. Nesse contexto, destacam-se ainda a instalação da Faculdade Filosofia Dom José em Sobral em 1961, a criação da Faculdade de Filosofia de Limoeiro do Norte seis anos depois (SULIANO, 2013, p. 43).

Quanto às universidades estaduais, em 1968, por iniciativa do governo municipal de Sobral, é criada a Universidade Estadual do Vale do Acaraú (UVA), tornada autarquia pelo Poder Executivo Estadual em 1984, e posteriormente fundação, em 1993. Anos depois é criada pela Lei Estadual nº 11.191 de 1986, a Universidade Regional do Cariri (URCA), até então Faculdade de Filosofia do Crato, oficialmente instalada no ano seguinte a promulgação da lei. Outra universidade estadual desse período é a UECE (Universidade Estadual do Ceará) (1975), cuja atuação através de sua rede multicampi foi de grande importância para ampliação do acesso ao ensino superior no Ceará (SULIANO, 2013). No ano de 2020 a UECE possui unidades de ensino atuando nos municípios de Itapipoca, Fortaleza, Guaiuba, Pacotí, Crateús, Quixadá,

Limoeiro do Norte, Tauá e Iguatu, que se voltam principalmente para a formação docente (UECE, 2020).

A despeito da contribuição da UECE, a instituição que teve protagonismo no referido processo de interiorização foi a UFC, principalmente a partir de 2006. Nesse ano, após a aprovação de sua participação no Programa de Expansão do Sistema Federal de Educação Superior, foram criados mais dois campi para a Universidade, no Cariri e em Sobral, e no ano seguinte foi instalado um terceiro campus, desta vez em Quixadá (SULIANO, 2013). Ao desmembrar-se da UFC em 2013 (Lei 12.826, de 5 de junho de 2013), o campus do Cariri tornou-se Universidade Federal do Cariri (UFCA), criando a possibilidade de nova ampliação de vagas para graduandos na região. Sediada em Juazeiro do Norte, a UFCA estende-se a outros quatro campi, localizados nos municípios de Barbalha, Crato, Brejo Santo e Icó (UFCA, 2020).

Destaca-se, por fim, que toda a atuação do Estado teve como objetivo principal expandir o ensino superior por meio da descentralização regional das universidades públicas. Porém, tal expansão realizou-se de modo que também tornasse possível “o desenvolvimento das regiões que apresentassem considerável potencial para a criação de polos socioeconômicos”, reconhecendo o papel da educação nesse processo (SULIANO, 2013, p. 53).

3 PROGRAMAS PARA O PRIMEIRO EMPREGO NO CEARÁ

Considerando a condição ‘não trabalha’ dos jovens ‘nem nem’, esta seção se dedica a descrição dos programas para o primeiro emprego desenvolvidos pelo Governo do Estado do Ceará, com foco na juventude. Nesse estado as principais iniciativas com vistas a resolução dessa questão social são o Projeto Primeiro Passo e o Programa E-Jovem. Ressalta-se a existência de outros programas como o Juventude Cidadã, o Criando Oportunidades e o Juventude Empreendedora, que se destinam ao mesmo fim e tem a mesma linha de ação do Primeiro Passo, ao objetivar primordialmente a preparação/capacitação do jovem cearense para tornar possível sua inserção no mercado de trabalho (SILVA, 2009; INESP, 2008).

O Primeiro Passo é uma ação de destaque do Governo do Estado direcionada à juventude cearense (SILVA; OLIVEIRA, 2014). Foi criado em 2007, durante o governo Cid Gomes, por meio da Secretaria do Trabalho e Desenvolvimento Social (STDS), tendo como entidade executora o Grupo de Apoio ao Investidor Social (GAIS) (GALVÃO, 2016).

São vários os seus objetivos específicos, que incluem, de modo geral, a criação de oportunidades voltadas à cidadania, inclusão social e profissional, ao promover aprendizado prático e experiências que possibilitam o desenvolvimento pessoal e profissional dos jovens atendidos (SILVA, 2009).

Quanto ao funcionamento, este se dá através de três modalidades distintas: Jovem Aprendiz, Jovem Estagiário e Jovem Bolsista. De modo geral podem participar jovens de comunidades vulneráveis, com idade entre 16 e 24 anos, que estejam cursando em escola pública o ensino fundamental II, educação especial, ensino médio ou que tenha concluído este último, e cuja renda familiar per capita seja igual ou inferior a $\frac{1}{2}$ salário mínimo. A participação do jovem em uma das modalidades será determinada tendo em vista sua idade e escolaridade, observando ainda os requisitos específicos de cada linha de ação do programa (SILVA, 2009).

Por meio do Jovem Aprendiz, o jovem com idade entre 16 e 24 anos, cursando o ensino médio, pode estagiar em empresas públicas ou privadas por um período de até 12 meses. Através do Jovem Bolsista, os adolescentes e jovens de 16 a 21 anos que cursam o ensino fundamental ou médio, tem acesso a ações de capacitação profissional e social com duração de 6 meses. Já o Jovem Estagiário destina-se a jovens do ensino médio de 16 a 21 anos e com

conhecimentos básicos em informática, levando-os a estagiar em empresas públicas ou privadas, durante 12 meses (INESP, 2008; SILVA, 2009).

Os recursos para o financiamento do projeto advêm do Fundo Estadual de Combate à Pobreza (FECOP), “tendo no período de 2007-2014, ocorrido investimentos na ordem de R\$ 58.175.031,25, e sido beneficiado em torno de 48 mil jovens”, nas três linhas de ação (SILVA; OLIVEIRA, 2014, p. 40).

Segundo Silva (2009, p. 84) esse é um dos projetos inovadores do governo cearense na área das políticas para a juventude, havendo, contudo, a necessidade de saber “se a capacitação oferecida e a prática profissional são, de fato, capazes de reduzir o desemprego juvenil, ou se a força de trabalho dos jovens é utilizada apenas como mão-de-obra barata para a substituição de profissionais adultos”. Os resultados de sua avaliação mostram que o programa conseguiu proporcionar ganhos sociais aos jovens participantes.

O E-Jovem também tem a preocupação com a transição entre a escola e o mundo do trabalho, mas age por meio da oferta de cursos técnicos na área de Tecnologia da Informação e Comunicação (TIC), utilizando-se de um ambiente virtual de aprendizagem. A iniciativa é do Governo do Estado do Ceará, por meio da Secretaria da Educação, e objetiva oferecer formação em Tecnologia da Informação e Comunicação (TIC) aos educandos do ensino fundamental, médio e egressos da rede pública estadual, com ênfase no protagonismo juvenil. São objetivos mais específicos a universalização da cultura digital, a qualificação profissional, a inserção no mundo do trabalho e o empreendedorismo social juvenil (TAHIM; SANTOS, 2012; FONTENELE, 2013).

A estratégia do projeto é fazer com que o jovem ou adolescente permaneça na escola ao mesmo tempo em que obtém uma formação mais sólida, que pode facilitar a sua entrada no mundo do trabalho. Para esse aprendizado é promovida uma formação em que o estudante desenvolve um processo autônomo de aprendizagem, passando a comprometer-se com sua própria formação (TAHIM; SANTOS, 2012) (FONTENELE, 2013).

A SEDUC destaca-se como “idealizadora e responsável por implementar, acompanhar e redirecionar as ações do projeto e integra o grupo gestor e normatizador” (p. 28), mas tem-se também como instituições principais na gestão do E-Jovem, o Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) e o Instituto Centro de Ensino Tecnológico (CENTEC). Além dessas, “outras instituições agregam valor ao projeto, com parcerias que incidem no repasse de metodologia e conteúdo específicos e na formação de educadores e instrutores em disciplinas específicas” (FONTENELE, 2013, p. 29).

Pode-se destacar ainda a respeito do E-Jovem, a importância da temática ‘mercado de trabalho’, tendo em vista que a formação proporcionada pelo projeto se propõe a qualificar os alunos para a atividade laboral e para a responsabilidade social (FONTENELE, 2013).

Apesar dos problemas identificados, a cada ano o projeto cresce, tendo suas metas ampliadas e expandindo-se entre os municípios cearenses. Como ressalta Fontenele (2013, p. 28), “desde a sua criação, o projeto já qualificou mais de 13.000 mil alunos em 180 escolas do estado de 70 municípios, mais de 40% de inserção no mercado de trabalho”. Nesse processo também cresce o volume de bolsas concedidas. Para atender esse quantum de estudantes, o investimento foi de aproximadamente 13 milhões de reais, considerando o período de 2008 a 2011.

Esta seção mostrou a atuação do Governo do Estado do Ceará com vistas a inserção do jovem cearense no mercado de trabalho. Os programas apresentados são relevantes em termos de público atendido, especialmente o Primeiro Passo, pelo qual já passaram mais de 48 mil jovens. O E-Jovem, apesar de menor, também contribui para amenizar o desemprego nesse

grupo, tendo qualificado mais de 13 mil jovens, além de conseguir inserir no mercado de trabalho boa parte dos participantes.

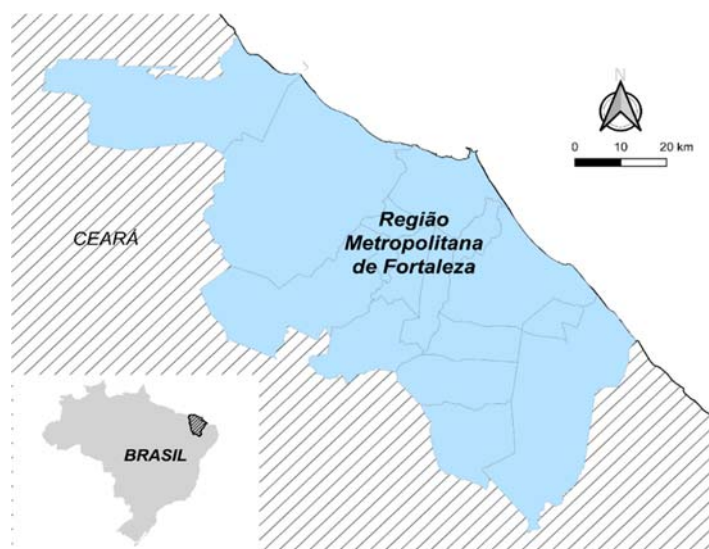
4 PROCEDIMENTOS METODOLÓGICOS

Para o alcance do principal objetivo deste estudo, esta seção apresenta detalhadamente os procedimentos metodológicos, ao trazer os esclarecimentos sobre a área de estudo, fonte e tratamento dos dados, e demais informações utilizadas.

O universo deste estudo contempla os jovens de 15 a 29 anos de idade, que não estudam e não trabalham (conhecidos na literatura como geração “nem nem”), residentes na Região Metropolitana de Fortaleza – RMF (Figura 1), no ano de 2015. Sendo assim, o Ceará, é o estado para os quais foram descritos os investimentos em educação e os programas para o primeiro emprego, dado que a metrópole cearense é o foco de análise sobre o perfil dos jovens ‘nem nem’.

Sobre essa metrópole, a mesma foi criada através da Lei Complementar Nº 14 de 8 de junho de 1973, que estabeleceu, ao todo, oito regiões metropolitanas no Brasil. A RMF foi instituída com 6 municípios e no ano de 2015 compõe-se de 15 municípios (Aquiraz, Cascavel, Caucaia, Chorozinho, Eusébio, Fortaleza, Guaiuba, Horizonte, Itaitinga, Maracanaú, Maranguape, Pacajus, Pacatuba, Pindoretama e São Gonçalo do Amarante).

FIGURA 1 : Mapa de localização da RMF



Fonte: Elaborado a partir da malha digital – IBGE (2010).

Quanto a fonte de dados para a realização do presente estudo, foram utilizados os microdados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2015, já que é possível desagregar até Regiões Metropolitanas. Para o tratamento estatístico dos dados foi utilizado o software SPSS (Statistical Package for the Social Science), e para obter os resultados do modelo econométrico o STATA 13. Em complemento foi utilizado o programa Excel para a construção de tabelas. Já a confecção do mapa de localização realizou-se com o QGis (versão 2.18.12).

Inicialmente a amostra foi dividida em três categorias de análise: (i) nem-nem (nem estuda e nem trabalha), (ii) somente trabalha e (iii) somente estuda, procurando dar a

dimensão/contingente dos jovens que não estudam nem trabalham em relação aos demais grupos.

Quanto ao perfil, as variáveis selecionadas na caracterização demográfica e socioeconômica da população entre 15 a 29 anos, que nem estuda e nem trabalha foram:

- I) Sexo: i) masculino e ii) feminino;
- II) Grupos etários: i) 15 a 19 anos, ii) 20 a 24 anos e iii) 25 a 29 anos;
- III) Raça/cor: i) branca, ii) preta, iii) amarela, iv) parda e v) indígena;
- IV) Nível de instrução: i) sem instrução e fundamental incompleto, ii) fundamental completo e médio incompleto, iii) médio completo e superior incompleto, iv) superior completo;
- V) Estado conjugal: i) solteiro(a), ii) casado(a), iii) desquitado(a)/separado(a) judicialmente, iv) divorciado(a) e v) viúvo(a);
- VI) Mora com a mãe: i) sim e ii) não;
- VII) Situação do domicílio: i) urbano, ii) rural;
- VIII) Rendimento domiciliar per capita em salários mínimos: i) até 1/2, ii) mais de 1/2 até 1, iii) mais de 1 até 2, iv) mais de 2 até 3, v) mais de 3 até 5, vi) mais de 5 até 10 e vii) acima de 10.

Já no modelo econométrico *probit* foi utilizada como variável dependente a *dummy* referente a ser nem nem (1) ou não (0), e como variáveis explicativas idade, escolaridade e as dummies sobre o sexo (0 para masculino e 1 para feminino) e o rendimento domiciliar per capita (0 para mais de 1/2 SM e 1 para até 1/2 SM). As variáveis faltantes em relação as estatísticas descritivas foram retiradas do modelo por não apresentarem significância estatística.

No intuito de calcular o sentido e o grau de influência das variáveis apresentadas anteriormente sobre a condição do jovem ser ou não nem nem, foi utilizado o modelo Probit. O mesmo método foi utilizado por Figueiredo e Almeida (2017) ao estudar a população nem nem no Brasil, a partir da PNAD de 2012.

O modelo escolhido baseia-se na utilização de uma variável dependente I_i determinada por uma ou mais variáveis explanatórias X_i . Através da variável dependente ou índice de utilidade observável é determinada a decisão da i -ésima observação do modelo, ou seja, quanto maior o valor do índice I_i , maior a probabilidade de uma decisão positiva para a observação. O índice I_i pode ser definido como:

$$I_i = \beta_1 + \beta_2 X_i \quad (1)$$

Tem-se que $Y = 1$ para indivíduos na condição “nem nem” e $Y = 0$ para caso contrário, considerando que o índice I_i possui um nível crítico ou limiar I_i^* que atua como um limite, tal que se I_i^* for menor ou igual a I_i o indivíduo pertencerá à condição “nem-nem”, e caso contrário não, como exposto em (2).

$$Y = \begin{cases} 1, & \text{se } I_i^* \leq I_i \\ 0, & \text{se } I_i^* > I_i \end{cases} \quad (2)$$

Seguindo a premissa de normalidade, a probabilidade de que I_i^* seja menor ou igual a I_i pode ser calculada a partir da função de distribuição ou densidade de probabilidade acumulada (FDA) normal padronizada, expressa em (3).

$$P_i = P(Y = 1|X) = P(I_i^* \leq I_i) = P(Z_i \leq \beta_1 + \beta_2 X_i) = F(\beta_1 + \beta_2 X_i) \quad (3)$$

em que $P(Y = 1|X)$ é a probabilidade do indivíduo estar na condição “nem-nem” dado um ou mais valores da variável explanatória X , e onde Z_i é a variável normal padronizada. F é a função de distribuição acumulada normal padrão, que é explicada em (4):

$$\begin{aligned} F(I_i) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X_i} e^{-z^2/2} dz \end{aligned} \quad (4)$$

Diferente dos parâmetros dos modelos de regressão linear comuns que fornecem informações sobre o sentido e a grandeza da influência de uma variável explicativa, os parâmetros do modelo *probit* indicam de forma imediata somente o sentido da influência, já que os valores dos parâmetros estimados significam a contribuição individual das variáveis no índice de utilidade que não é observável, mas que está relacionado com a probabilidade do evento ocorrer.

Assim, informações sobre a grandeza da influência somente são obtidas através de cálculos adicionais. Como no modelo *probit* todos os regressores estão relacionados no cálculo da variação da probabilidade, a taxa desta variação, ou seja, a grandeza da influência será dada por $\beta_j f(Z_i)$, onde $f(Z_i)$ é a função de densidade da variável normal padrão e $Z_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$, isto é, é o modelo de regressão utilizado na análise.

Segundo Oliveira (1998), o pseudo- R^2 obtido no modelo probit, diferente do R^2 da regressão clássica não tem interpretação intuitiva imediata, de modo que um valor muito próximo de 1 indicaria mais uma má especificação do que um bom ajustamento. Dessa forma, o comando ‘*estat class*’ foi considerado como elemento de apreciação da qualidade dos resultados estimados, por informar o percentual de acertos do modelo. Tal comando consiste de uma matriz na qual se confrontam as previsões do modelo, com a repartição efetiva das observações da amostra entre os dois grupos, associados ao 0 e ao 1. Ao calcular para cada observação uma estimativa da probabilidade de $Y = 1$ (por meio das fórmulas apresentadas anteriormente), é predito que $Y_i = 1$ ou $Y_i = 0$, conforme essa estimativa exceda ou não 0,5. Em seguida a informação é organizada numa tabela da seguinte forma:

	Y _i predito = 1	Y _i predito = 0	Totais
Y _i observado = 1	A	B	n1
Y _i observado = 0	C	D	n0
Totais	E	F	n

A e D representam os números correspondentes as previsões acertadas do modelo, enquanto B e C quantificam as previsões erradas. Assim, $(A+D)/n$ mostra o percentual de acertos do modelo, ao passo que a proporção de observações erradamente classificadas, $(B+C)/n$, é a chamada taxa aparente de erro.

5 RESULTADOS

5.1 Quantos são?

A presente seção volta-se para análise do perfil do jovem nem nem na RMF. Inicialmente, a Tabela 1 mostra o volume de jovens que não estudam, não trabalham ou apresentam ambas as características, bem como suas respectivas participações no total de jovens na metrópole cearense.

Inicialmente constata-se que a RMF concentra mais de um milhão (1.021.543) de jovens com idade de 15 a 29 anos. Sobre aqueles que não estudam, a metrópole registra em termos absolutos (657.802) e relativos (64%) a maioria da sua população jovem nessa condição. Quanto aos que não trabalham, 48,81% estão fora do mercado de trabalho. Com relação aos jovens que não estudam e não trabalham (nem nem), população alvo desse estudo, coube 23,87%, sendo um contingente expressivo em condição de vulnerabilidade social e econômica, com forte impacto nas suas vidas e no crescimento e desenvolvimento da RMF, bem como do Ceará e do Brasil.

TABELA 1 : Volume e participação dos jovens que nem estudam e nem trabalham ('nem nem') no total de jovens de 15 a 29 anos de idade – RMF - 2015

Condição do Jovem	Número de Jovens	Part. (%) no Total da RMF*
Nem Nem	243.811	23,87
Não Estudam	657.802	64,39
Não Trabalham	498.572	48,81
Pop. de 15 a 29 anos	1.021.543	-

*Condições/categorias não são excludentes, em razão disso os percentuais não somam 100%.

Fonte: Elaboração própria a partir dos microdados da PNAD 2015.

Na Tabela 2 é possível ver os dados anteriores desagregados segundo os grupos etários de 15 a 19, 20 a 24 e 25 a 29 anos. A distribuição entre estes segundo a população total não apresenta grande variação, visto que cabe a cada grupo etário aproximadamente um terço do total de jovens, sendo que aqueles de 15 a 19 anos estão em segunda posição, entretanto, é o grupo em que há menos jovens na condição de nem nem (26,56%).

Essa constatação pode ser reflexo dos avanços obtidos na educação cearense, resultante, entre outros objetivos, da busca pela inserção, cada vez mais, de crianças e adolescentes na rede de ensino, como pôde ser visto nas seções anteriores, destacando-se o maior esforço em relação aos jovens do ensino médio, associado a identificação de problemas recorrentes nesse nível de ensino.

TABELA 2 : Volume e participação dos jovens que nem estudam e nem trabalham ('nem nem') no total de jovens de 15 a 29 anos de idade, segundo o grupo de idade - RMF - 2015

Grupo Etário	População Total		Não Estudam		Não Trabalham		Nem Nem	
	Abs.	(%)	Abs.	(%)	Abs.	(%)	Abs.	(%)
15 a 19	343.575	33,63	112.878	17,16	262.901	52,73	64.755	26,56
20 a 24	347.120	33,98	263.970	40,13	142.249	28,53	101.559	41,65
25 a 29	330.848	32,39	280.954	42,71	93.422	18,74	77.497	31,79
15 a 29	1.021.543	100,00	657.802	100,00	498.572	100,00	243.811	100,00
Part. (%) no total da RMF*	-		64,39		48,81		23,87	

*Condições/categorias não são excludentes, em razão disso os percentuais não somam 100%.

Fonte: Elaboração própria a partir dos microdados da PNAD 2015.

A divisão por grupo etário é conveniente, pois o volume de jovens que não estuda, não trabalha ou ambos, está grandemente relacionado a idade em que se encontra o jovem. Como pode ser visto, a medida em que aumenta a idade, cresce o percentual de jovens que não estudam e diminui os que não trabalham. Em razão disso, a maior participação dos nem nem é observada no grupo de idade intermediário (20 a 24 anos), sendo este o momento de transição entre o mundo da escola e o do trabalho.

5.2 Quem São?

Considerando o perfil demográfico quanto ao sexo dos jovens nem nem de 15 a 29 anos, observa-se que a maioria são mulheres, com participações acima de 60% (Tabela 3). O resultado encontrado por Camarano et al. (2006) para os anos de 1980 e 2010 mostra que as mulheres também são maioria nessa condição no Brasil, apesar de sua participação ter diminuído. Associando os resultados ao fenômeno de transição para a vida adulta, sugerem que para as mulheres não há necessariamente uma passagem pelo mercado de trabalho e possivelmente elas saem do domicílio de origem como cônjuges ou mães, sem ter realizado alguma atividade econômica. O mesmo resultado é confirmado por outros estudos como os de Figueiredo e Almeida (2017) e Remy e Vaz (2017), cujos resultados indicam forte efeito de gênero no fenômeno dos nem nem.

TABELA 3 : Sexo dos jovens ‘nem nem’, por grupo etário - RMF - 2015

Sexo	Jovens Nem Nem			15 a 29
	15 a 19	20 a 24	25 a 29	
Masculino	40,44	36,58	24,20	33,67
Feminino	59,56	63,42	75,80	66,33
Total	100,00	100,00	100,00	100,00

Fonte: Elaboração própria a partir dos microdados da PNAD 2015.

A respeito da divisão entre grupos etários, chama atenção a diferença entre homens e mulheres na evolução dos percentuais. No caso deles, idades mais elevadas correspondem a participações menores no total de jovens na condição analisada, possivelmente associadas ao motivo de trabalho, enquanto para as mulheres ocorre o inverso.

A maior participação feminina é registrada no grupo etário de 25 a 29 anos, onde 75,80% dos nem nem são mulheres. Por conseguinte, também tem-se aí a menor participação masculina (24,20%). Isto pode estar relacionado a deixar ou não dar continuidade aos estudos, mas também a dificuldades para conseguir trabalho, ou escolhas na vida pessoal. Ademais, a diferença entre homens e mulheres quanto a transição entre adolescência e vida adulta é um fator importante a ser considerado, dado a predominância feminina na condição de nem nem (CARDOSO, 2013).

Por outro lado, há os casos em que para as mulheres o abandono da vida escolar e a inatividade econômica se devem a frustrações/dificuldades, como no caso de gravidez na adolescência, correspondente a uma gravidez indesejada entre as mais jovens, que traz a necessidade de cuidar do filho ou em outros casos cuidar de irmãos mais novos ou pais idosos. Como destacado por Remy e Vaz (2017, p. 136): “o cuidado com os filhos, ao recair primordialmente sobre as mulheres, penaliza sua escolaridade e inserção no mercado de trabalho, tornando-as mais propensas à condição estudada”.

Outra característica que interfere tanto na vida estudantil quanto na inserção no mercado de trabalho é a raça/cor. Na Tabela 4 é possível visualizar que independente do grupo

etário, os jovens nem nem são na grande maioria de raça/cor parda. Em segundo lugar tem-se a raça/cor branca. Amarelos e indígenas apresentam participações quase sempre próximas a zero.

TABELA 4 : Raça/cor dos jovens ‘nem nem’, por grupo etário - RMF - 2015

Raça/Cor	Jovens Nem Nem			15 a 29
	15 a 19	20 a 24	25 a 29	
Branca	25,13	23,70	21,01	23,22
Preta	3,83	4,53	6,85	5,08
Amarela	1,09	0,00	0,00	0,29
Parda	69,40	71,77	72,14	71,26
Indígena	0,55	0,00	0,00	0,15
Total	100,00	100,00	100,00	100,00

Fonte: Elaboração própria a partir dos microdados da PNAD 2015.

Conforme dados do Censo Demográfico 2010, no Ceará, assim como no Nordeste, se sobressai o número de residentes pardos, o que implica em maior percentual de jovens com essa raça/cor (IPECE, 2012). A segunda maior participação coube aos autodeclarados brancos, que representam 23,22% dos jovens nem nem da RMF. A análise para o grupo etário geral (15 a 29 anos) é representativa em relação aos grupos etários específicos, visto que, em geral, os valores relativos não mostram grandes variações entre eles.

Entretanto, é preciso destacar que, no grupo de idade entre 25 a 29 anos, há maior percentual de negros em relação ao de menor idade (15 a 19). A diferença mesmo que pequena, pode ser um indicativo de que a discriminação por raça/cor está mais presente no mercado de trabalho, diminuindo as chances de conseguir emprego. Por outro lado, mesmo após a entrada no mercado de trabalho, esta discriminação pode ser constatada através de diferenças salariais, e não somente em relação à raça/cor, mas também por sexo (CACCIAMALI; HIRATA, 2005).

Quanto ao nível de instrução, a Tabela 5 mostra que aproximadamente metade dos jovens nem nem tem no mínimo o ensino médio completo, o que não é um bom resultado, pois indica dificuldades de continuar se qualificando ou de inserção no mercado de trabalho. Na RMF aproximadamente 47% dos jovens que não estudam e não trabalham são escolarizados até o ensino médio completo ou superior incompleto. Com a segunda maior participação aqueles que não têm instrução ou não terminaram o ensino fundamental, correspondentes a aproximadamente 28%. Em seguida tem-se o percentual de jovens com fundamental completo ou médio incompleto (22,21%), enquanto a menor participação coube aos nem nem com superior completo (3,34%).

TABELA 5 : Nível de instrução dos jovens ‘nem nem’, por grupo etário - RMF - 2015

Nível de Instrução	Jovens Nem Nem			15 a 29
	15 a 19	20 a 24	25 a 29	
Sem Instrução e Fundamental Incompleto	35,52	25,78	23,29	27,58
Fundamental Completo e Médio Incompleto	24,04	21,95	21,01	22,21
Médio Completo e Superior Incompleto	40,43	48,78	49,77	46,88
Superior Completo	0,00	3,48	5,93	3,34
Total	100,00	100,00	100,00	100,00

Fonte: Elaboração própria a partir dos microdados da PNAD 2015.

Em geral, foi possível observar que os dois níveis de instrução mais baixos se concentram no grupo etário mais jovem (15 a 19), ao passo que a maior escolarização é verificada no grupo mais velho (25 a 29). Uma maior escolaridade significa um capital humano ocioso. Por outro lado, a menor escolaridade para o grupo etário de 15 a 19 anos pode contribuir para dificultar a inserção, já que indica o encerramento ou abandono dos estudos muito cedo. Assim, as ocupações quando encontradas por esses jovens são precárias ou em atividades de curta duração, levando a elevadas taxas de rotatividade e desemprego (MONTEIRO, 2013).

A análise do estado civil dos jovens nem nem (15 a 29 anos) mostra que a maioria é solteira, com percentuais frequentemente acima de 90%, independente do grupo etário que façam parte (Tabela 6). Essa questão pode estar relacionada, segundo Barbosa e Rocha-Coutinho (2007, p.168), a "[...] uma opção de vida que envolve diferentes questões, como encontrar o parceiro certo, investir em uma carreira profissional, conquistar estabilidade financeira e efetiva, manter a liberdade, entre outros".

TABELA 6 : Estado civil dos jovens ‘nem nem’, por grupo etário - RMF - 2015

Estado Civil	Jovens Nem Nem			15 a 29
	15 a 19	20 a 24	25 a 29	
Solteiro (a)	99,45	95,09	88,10	94,46
Casado (a)	0,00	4,15	7,73	3,91
Desquit. (a) ou separado (a) judicialmente	0,55	0,38	0,60	0,49
Divorciado (a)	0,00	0,38	2,98	0,98
Viúvo (a)	0,00	0,00	0,59	0,16
Total	100,00	100,00	100,00	100,00

Fonte: Elaboração própria a partir dos microdados da PNAD 2015.

Quanto aos jovens nem nem que moram com a mãe ou não, os dados da Tabela 7 mostram percentuais mais expressivos para jovens que tem a mãe morando na residência. A respeito da subdivisão entre os grupos etários, observa-se clara relação entre o aumento da idade e a saída da casa dos pais, isto é, embora no grupo geral a maioria dos nem nem more com a mãe, isto se deve aos dois grupos etários mais jovens, especialmente o de 15 a 19 anos, já que entre 25 e 29 quase 60% não mora com a mesma. Contudo, mesmo nesse grupo o percentual daqueles que moram com a mãe na RMF é consideravelmente elevado (40,41%), o que é esperado em se tratando dos jovens nem nem, pois devido a condição de não trabalhar, torna-se difícil estabelecer e manter um domicílio.

TABELA 7 : Jovens ‘nem nem’ que moram com a mãe ou não, por grupo etário - RMF - 2015

Mora com a Mãe	Jovens Nem Nem			15 a 29
	15 a 19	20 a 24	25 a 29	
Mora	72,83	54,38	40,41	55,04
Não mora	27,17	45,62	59,59	44,96
Total	100,00	100,00	100,00	100,00

Fonte: Elaboração própria a partir dos microdados da PNAD 2015.

O elevado percentual de jovens morando com os pais está de acordo com a literatura sobre os nem nem e a chamada geração canguru, ao destacarem a permanência dos jovens junto da família, devido ao conforto proporcionado ou a falta de perspectivas no mercado de trabalho

que os leva a inatividade econômica, por vezes associada a inatividade escolar (REMY; VAZ, 2017).

Em se tratando da situação de domicílio dos jovens nem nem, a grande maioria (cerca de 94%) reside em domicílio urbano, como mostra a Tabela 8. Na RMF, dentre aqueles com idade entre 15 e 29 anos, 93,61% residem no urbano e apenas 6,39% no rural, havendo pouca diferenciação entre os grupos etários. Os percentuais encontrados para o urbano foram elevados devido a área estudada se tratar de uma metrópole, que são áreas muito urbanizadas.

TABELA 8 : Situação do domicílio dos jovens ‘nem nem’, por grupo etário - RMF - 2015

Situação do Domicílio	Jovens Nem Nem			15 a 29
	15 a 19	20 a 24	25 a 29	
Urbano	93,99	93,73	93,15	93,61
Rural	6,01	6,27	6,85	6,39
Total	100,00	100,00	100,00	100,00

Fonte: Elaboração própria a partir dos microdados da PNAD 2015.

Tal resultado pode ser explicado em função da atração que o urbano exerce sobre os indivíduos, que o veem como um espaço de oportunidades. Assim, além dos jovens já residentes no meio urbano, há a migração do rural para tal área, em busca de estudo, trabalho e/ou outros aspectos de qualidade de vida (HARTWIG, 2012), muitas vezes tendo abandonado ou concluído os estudos, tendo em vista sua situação econômica (GARCIA et al., 2012).

Entretanto, há nos centros urbanos grande seletividade migratória (QUEIROZ, 2013) e rotatividade de trabalhadores que elevam o desemprego, principalmente entre os mais jovens, enquanto no meio rural a menor escolaridade e o menor dinamismo econômico inibem o movimento de oferta e demanda de trabalho, respectivamente, contribuindo para a permanência do trabalhador no emprego por mais tempo (MONTE; PENIDO, 2008).

Como indicado acima por Garcia et al. (2012), os jovens brasileiros estão em famílias com renda baixa, tendo em vista que 83% apresentam renda domiciliar per capita inferior a dois salários mínimos. A Tabela 9 traz estas informações mostrando o perfil do jovem nem nem quanto ao rendimento domiciliar per capita.

TABELA 9 : Faixas de rendimento domiciliar per capita dos jovens ‘nem nem’, em salários mínimos, por grupo etário - RMF – 2015

Rendimento Domiciliar Per Capita	Jovens Nem Nem			15 a 29
	15 a 19	20 a 24	25 a 29	
Até 1/2 SM	52,51	64,51	58,88	59,52
Mais de 1/2 até 1 SM	39,10	28,32	27,57	30,95
Mais de 1 até 2 SM	7,82	6,09	8,88	7,44
Mais de 2 até 3 SM	0,00	0,36	3,74	1,34
Mais de 3 até 5 SM	0,00	0,72	0,93	0,60
Mais de 5 SM	0,56	0,00	0,00	0,15
Total	100,00	100,00	100,00	100,00

Fonte: Elaboração própria a partir dos microdados da PNAD 2015.

Na RMF aproximadamente 60% possui rendimento domiciliar per capita abaixo de meio salário mínimo, com maior peso no grupo de idade de 20 a 24 anos (64,51%). Com isso, cerca de 31% residem em domicílio com rendimento entre meio e 1 salário. Por outro lado,

somente 7,44% tem rendimento entre 1 e 2 salários mínimos, e as demais categorias apresentam valores pequenos, frequentemente inferiores a 1%.

Na seção sobre os programas para o primeiro emprego de jovens, havia aqueles que usavam o baixo rendimento como um critério para delimitar o público a ser atendido pela iniciativa. Nesse contexto, foi destacado que tais critérios poderiam levar ao atendimento de menos jovens, quando a realidade do desemprego juvenil os atingia de maneira geral. Em relação ao Programa Nacional de Estímulo ao Primeiro Emprego para os Jovens (PNPE), Weyrich (2007) destacou a necessidade de ações estatais voltadas para todos os jovens. Entretanto, os dados analisados para os jovens nem nem da RMF confirmam que a maioria se encontra em famílias com baixo rendimento, sendo importante políticas públicas de estímulo ao emprego e educação, notadamente para os jovens de baixa renda, que são a grande maioria.

5.3 Resultados do modelo econométrico

Com a aplicação do modelo *probit* foi possível obter informações referentes ao sentido da influência de cada variável na probabilidade dos indivíduos observados estarem na condição de nem nem. Dentre as variáveis utilizadas nas estatísticas descritivas, aquelas referentes a raça/cor, estado civil, morar com a mãe e situação de domicílio não foram significantes, e por isso foram retiradas do modelo. As informações estatísticas referentes aos dados utilizados indicaram como resultado do comando '*estat class*', um percentual de acertos de 79% para a RMF. Ademais, o resultado para o sentido da influência mostra que apenas a escolaridade afeta negativamente a probabilidade de ser nem nem, enquanto 'Feminino', 'Idade' e 'AtémioSM' tem efeito positivo (Tabela 10).

TABELA 10 : Resultados do modelo *probit* - RMF – 2015

Nem Nem	Coefficiente	Erro Padrão	z	P > z
Feminino	0,5411395	0,0563	9,61	0,000
Idade	0,0459134	0,0067	6,88	0,000
Escolaridade	-0,0616778	0,0097	-6,36	0,000
AtémioSM	0,7405131	0,0582	12,72	0,000
Intercepto	-1,730888	0,1691	-10,24	0,000

Fonte: Resultados do modelo *probit*, a partir dos microdados da PNAD 2015.

Além do sentido da influência foi estimado o efeito marginal associado a probabilidade do evento '1' (ser nem nem) ocorrer. As estimativas mostram que ser do sexo feminino aumenta em aproximadamente 15,78% a probabilidade de ser nem nem. Para a idade foi considerado o intervalo de 15 a 29 anos, com isso, cada ano a mais eleva a probabilidade de estar na condição de nem nem em, aproximadamente, 1,35%. O efeito positivo deriva dos maiores percentuais de jovens nessa condição nos dois últimos grupos etários (20 a 24 e 25 a 29).

TABELA 11: Efeito marginal das variáveis – RMF – 2015

Variável	dy/dx	Erro Padrão	z	P > z	X
Feminino*	0,1577614	0,0161	9,82	0,000	0,50830
Idade	0,0134957	0,0020	6,87	0,000	21,9408
Escolaridade	-0,0181295	0,0029	-6,36	0,000	9,70975
AtémioSM*	0,2320184	0,0188	12,31	0,000	0,35885

$y = \text{Pr}(\text{nemnem}) (\text{predict}) = 0,21722603; 0,27257462; 0,2093328$, respectivamente.

(*) dy/dx é para a mudança discreta da variável *dummy* de 0 para 1

Fonte: Resultados do modelo *probit*, a partir dos microdados da PNAD 2015.

Quanto a escolaridade, única variável a apresentar influência negativa sobre o evento analisado, foi observado que cada ano a mais de instrução reduz a probabilidade de ser classificado como nem nem em 1,81%.

A *dummy* associada ao rendimento familiar, por sua vez, mostrou que possuir rendimento per capita abaixo de meio salário mínimo eleva a probabilidade do jovem ser nem nem em cerca de 23%, sendo, portanto, a variável explicativa com maior impacto, reafirmando os resultados destacados nas estatísticas descritivas.

6 CONSIDERAÇÕES FINAIS

Dado o significativo crescimento no contingente de jovens simultaneamente fora da escola e do mercado de trabalho, estudos sobre a geração nem nem (nem estuda e nem trabalha) tornaram-se foco de análise na literatura brasileira, constituindo-se um tema de grande relevância social e, portanto, de interesse de pesquisadores e formuladores de políticas públicas. Diante disso, o objetivo principal deste estudo foi saber quantos são e quem são os jovens nem nem residentes na Região Metropolitana de Fortaleza, dado a escassez de estudos sobre os mesmos nessa metrópole nordestina.

O estudo inicialmente buscou descrever os investimentos realizados em educação no Ceará, dado que essa metrópole se encontra nesse estado. A revisão da literatura mostrou o maior esforço na educação infantil, com programas mais notáveis para este nível de ensino (PAIC, Mais PAIC, Mais Educação), enquanto o programa para o ensino médio de maior atuação derivava da iniciativa federal, e no ensino superior verifica-se implantação tardia e desenvolvimento lento, apesar da interiorização.

Levando em consideração a condição ‘não trabalha’, este estudo descreveu os programas para o primeiro emprego desenvolvidos pelo Governo do Ceará, dada a problemática do desemprego juvenil. Os programas estaduais destinam-se, de modo geral, a adolescentes e jovens desde os 14 aos 29 anos, mas cada um com condicionalidades específicas em relação a idade, escolaridade etc. Em geral, tais programas atuam sobre a problemática através da capacitação, qualificação ou profissionalização, constituindo-se em ações educacionais e/ou sociais. O intuito é inserir no mercado de trabalho os jovens atendidos, mas impõe como desafio a contratação em si não depender do governo estadual, que atua indiretamente no incentivo a geração de postos de trabalho destinados aos jovens e na intermediação e/ou estímulo a contratação destes.

Como principal contribuição deste estudo, a análise dos resultados sobre os jovens nem nem na RMF, os caracterizou como maioria do sexo feminino, com predominância da raça/cor parda e branca, com cerca de 50% com ensino médio completo ou superior incompleto, apesar do elevado percentual de pessoas com escolaridade inferior a esta (cerca de 45%), majoritariamente solteiros, residindo com os pais, em domicílio urbano, e com baixo rendimento familiar.

Ademais foi utilizado o modelo *probit* para verificar a influência de algumas dessas características sobre a probabilidade do jovem ser nem nem na RMF. As estimativas revelam como variável de maior impacto o rendimento per capita abaixo de meio salário mínimo, seguido da influência positiva em ser do sexo feminino, e reafirmou a importância da escolaridade na redução de tal probabilidade. Assim, os resultados para a RMF estão em acordo com pesquisas já realizadas sobre a temática, destacando a predominância da participação feminina e o baixo rendimento domiciliar entre os jovens que não estudam e não trabalham.

Por fim, convém reafirmar a relevância da temática abordada, que envolve o âmbito social, por se tratar das perspectivas futuras dos jovens, e econômico, devido a importância desse segmento para o desenvolvimento do país. Assim, torna-se importante saber mais a respeito e, principalmente, agir por meio da continuidade e aperfeiçoamento de iniciativas para a melhora na educação, qualificação profissional e geração de empregos, pois se não fossem essas medidas e programas, a situação dos jovens na RMF seria ainda mais preocupante. Com isso, além da contribuição acadêmica, espera-se que os resultados deste estudo auxiliem na focalização de políticas públicas para o combate à evasão escolar e ao desemprego, ou ainda, para a saída da condição de nem nem.

7 REFERÊNCIAS

BARBOSA, Patrícia Zulato; ROCHA-COUTINHO, Maria Lúcia. Maternidade: novas possibilidades, antigas visões. *Psicologia Clínica*, Rio de Janeiro, v. 19, p. 163/1-185, 2007.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. **Saúde na escola** / Ministério da Saúde, Secretaria de Atenção à Saúde, Departamento de Atenção Básica. - Brasília: Ministério da Saúde, 2009. 96 p.: il. - (Série B. Textos Básicos de Saúde) (Cadernos de Atenção Básica; n. 24).

CACCIAMALI, Maria Cristina; HIRATA, Guilherme Issamu. A Influência da Raça e do Gênero nas Oportunidades de Obtenção de Renda - Uma Análise da Discriminação em Mercados de Trabalho Distintos: Bahia e São Paulo. **Estudos Econômicos** (São Paulo. Impresso), São Paulo, v. 35, n.4, p. 767-795, 2005.

CAMARANO, Ana Amélia; KANSO, Solange; MELLO, Juliana Leitão; ANDRADE, Adriana. Estão fazendo a transição os jovens que não estudam, não trabalham e não procuram trabalho? In: **Transição para a vida adulta ou vida adulta em transição?** 1ed. Rio de Janeiro: IPEA, 2006, v. 1, p. 259-290.

CARDOSO, Adalberto. Juventude, trabalho e desenvolvimento: elementos para uma agenda de investigação. **Caderno CRH** (UFBA. Impresso), Salvador, v. 26, n. 68, p. 293-314, 2013.

CEARÁ. Secretaria da Educação (SEDUC - CE). **Projetos e programas**. Disponível em: < <http://www.seduc.ce.gov.br/index.php/projetos-e-programas> >. Acesso em: 8 de out. de 2016.

_____. Secretaria da Educação (SEDUC - CE). **SPAECE - Sistema Permanente de Avaliação da Educação Básica**. Disponível em: < <https://www.seduc.ce.gov.br/spaace/> >. Acesso em: 03 de junho de 2020.

COELHO, Maria Izolda Cela de Arruda. **Rede de cooperação entre escolas: uma ação no âmbito do programa alfabetização na idade certa - PAIC**. Dissertação (Mestrado Profissional em Gestão e Avaliação da Educação Pública) - Centro de Políticas Públicas e Avaliação da Educação. Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

COSTA, Joana Simões de Melo; ULYSSEA, Gabriel. O fenômeno dos jovens nem-nem. In: **Desafios à trajetória profissional dos jovens brasileiros**. 1ed. Rio de Janeiro: Ipea, 2014, v. 1, p. 115-140.

FIGUEIREDO, A. M. R.; ALMEIDA, J. B. S. A. **População Nem-Nem: uma análise a partir dos dados da PNAD 2012**. *Revista de Estudos Sociais*, v. 19, p. 106-129, 2017. Disponível

em: < <http://periodicoscientificos.ufmt.br/ojs/index.php/res/article/view/4942/pdf> >. Acesso em: 23 de maio de 2018.

FONTENELE, Paulo Sérgio. **O Projeto E-Jovem no Estado do Ceará: Perspectivas, Análise e Desafios**. Dissertação (Mestrado Profissional em Gestão e Avaliação da Educação Pública) - Centro de Políticas Públicas e Avaliação da Educação. Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

FUNDAÇÃO PERSEU ABRAMO (FPA). **Os intocáveis (V): a saga dos jovens brasileiros excluídos do trabalho e da educação**. FPA Comunica n.11. São Paulo, nov. 2013.

GALVÃO, Thais Leite. **Retrato do desemprego juvenil no Brasil e no Ceará nos anos de 2004 e 2014**. Monografia (Graduação em Ciências Econômicas). Universidade Regional do Cariri, Crato/CE, 2016.

GARCIA, Maria de Fátima; ARAÚJO, Eliane Cristina de; ARAÚJO, Elisângela Luzia; FAUSTINO, Izabel Aparecida. A Condição do Jovem no Mercado de Trabalho Brasileiro: uma análise comparativa entre o emprego e o primeiro emprego (1999-2009). **Revista ANPEC**, Brasília (DF), v. 13, p. 481-506, set/dez 2012.

HARTWIG, Marisa. **Migração campo cidade: trajetórias de vida, trabalho e escolarização de jovens trabalhadores**. In: I Seminário Internacional e I Fórum de Educação do Campo da Região Sul do RS: Campo e cidade em busca de caminhos comuns, Pelotas/RS, 2012.

IBGE. Instituto Brasileiro de Geografia e Estatística. **PNAD (2015)**. Microdados da Pesquisa Nacional por Amostra de Domicílio. Rio de Janeiro, 2016.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **IDEB: Resultados e Metas**. Disponível em: < <http://ideb.inep.gov.br/> >. Acesso em: 03 de junho de 2020.

INESP. Instituto de Estudos e Pesquisas sobre o Desenvolvimento do Estado do Ceará. **Inclusão Social e Formação Profissional da Juventude Cearense: Linhas de Ação, Programas e Projetos**. INESP: Fortaleza - CE, 2008.

IPECE. Instituto de Pesquisa e Estratégia Econômica do Ceará. Informe nº. 23. **Perfil da raça da população cearense**. Fortaleza - CE, março, 2012.

LIMA, Alessio Costa. Ciclo de avaliação da educação básica do Ceará: principais resultados. **Estudos em Avaliação Educacional**, São Paulo, v. 23, n. 53, p. 38-58, set/dez. 2012.

MONTE, Paulo Aguiar do; PENIDO, Mariângela Rangel Jacques. Determinantes da duração esperada do emprego urbano e rural no Nordeste brasileiro. **Revista de Economia e Sociologia Rural** (Impresso), Piracicaba-SP, vol. 46, n 04, p. 989-1014, out/dez 2008.

MONTEIRO, Joana da Costa Martins. Quem são os Jovens Nem-Nem? Uma análise sobre os jovens que não estudam e não participam do mercado de trabalho. In: **Ensaio IBRE de economia brasileira I**. 1ed. Rio de Janeiro: FGV, 2013, v. 1, p. 231-248.

NASPOLINI, Antenor. A reforma da educação básica no Ceará. **Estudos Avançados**. São Paulo, vol.15, n.42, May/Aug. 2001.

OLIVEIRA, M. M. **Modelos de escolha binária**. Faculdade de Economia do Porto, Portugal. 1998. Disponível em: <

https://www.fep.up.pt/disciplinas/2E103/modelos_de_escolha_binaria.pdf>. Acesso em: 19 de maio de 2018.

PAIC. Programa de Alfabetização na Idade Certa. **MAIS PAIC**: história. Disponível em: < <https://paic.seduc.ce.gov.br/index.php/o-paic/historico/historia> >. Acesso em: 03 de junho de 2020.

QUEIROZ, Silvana Nunes de. **Migrações, Retorno e Seletividade no Mercado de Trabalho Cearense**. Tese (Doutorado em Demografia) - Universidade Estadual de Campinas, Campinas - SP, 2013.

REMY, Maria Alice Pestana de Aguiar; VAZ, Daniela Verzola. Fora da escola e do mercado de trabalho: o jovem “nem-nem” no Estado do Rio de Janeiro. **Revista da ABET**, v. 16, n. 2, Jul./Dez. de 2017.

SANTIAGO, Lindelvania Matias de; RODRIGUES, Malvina Thaís Pacheco; OLIVEIRA JUNIOR, Aldivan Dias de; MOREIRA, Thereza Maria Magalhães. Implantação do Programa Saúde na Escola em Fortaleza - CE: atuação de equipe da Estratégia Saúde da Família. **Revista Brasileira de Enfermagem**, Brasília, vol. 65, nº 6, Nov./Dez. 2012.

SILVA, Joaquim Leitão de Castro e; OLIVEIRA, Lafaete Almeida de. **Juventude, Trabalho e Políticas Públicas**: Um Olhar sobre o Projeto Primeiro Passo do Governo do Estado do Ceará desenvolvido em Fortaleza - CE. Trabalho de Conclusão de Curso (Curso de Especialização em Gestão Pública) - Universidade da Integração Internacional da Lusofonia Afro-brasileira, Redenção/CE, 2014.

SILVA, Liduína Elisabeth Angelim Gomes da. **Trabalho e juventude**: uma análise do Projeto Primeiro Passo/CE. Dissertação (Mestrado Profissional em Planejamento e Políticas Públicas) - Universidade Estadual do Ceará, Fortaleza, 2009. 120 p.

SULIANO, Daniel Cirilo. **Expansão e interiorização da Universidade Federal do Ceará (UFC)**: avaliação das repercussões educacionais e sociais. Dissertação (Mestrado Profissional em Políticas Públicas e Gestão da Educação Superior) - Universidade Federal do Ceará, Fortaleza, 2013.

TAHIM, Demetrius Oliveira; SANTOS, Maria Inês Detsi de Andrade. **Considerações sobre o Projeto E-Jovem (SEDUC/CE) a partir dos Conceitos de Broadcast e Interatividade**. In: I Simpósio Internacional de Educação a Distância, São Carlos, 2012. Anais do SIED: ENPED, 2012.

UECE. Universidade Estadual do Ceará. **UECE**: Onde estamos. Disponível em: < <http://www.uece.br/onde-estamos/> >. Acesso em: 02 de junho de 2020.

UFCA. Universidade Federal do Cariri. **Instituição**: apresentação e história. Disponível em: < <https://www.ufca.edu.br/instituicao/apresentacao-e-historia/> >. Acesso em: 02 de jun. de 2020.

WEYRICH, Márcio Luiz. **Primeiro emprego: Uma análise comparativa dos programas federal e estaduais**. Monografia (Curso de Especialização em Orçamento Público) – Centro de Formação, Treinamento e Aperfeiçoamento (Cefor) da Câmara dos Deputados, Biblioteca Digital da Câmara dos Deputados. Brasília, 2007.

