

Essays on Environmental Economics and Human Capital

by

Juliana Cunha Carneiro Pinto

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Department of Economics
Birmingham Business School
College of Social Sciences
University of Birmingham
December 2020

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Chapter 1 - We quantify the impact of two educational policies in Brazilian public schools on (i) scores in the national examination *Prova Brasil* and (ii) the dropout and pass rates. We construct a unique longitudinal panel at the school level from 2007 to 2015, combining administrative data from several government sources. To identify the causal impacts stemming from the two policies we use Propensity Score Matching with Difference-in-differences (PSM-DID). The results reveal a positive impact of the school entrance age policy on Portuguese and Mathematics scores and on dropout rates; and a negative effect on pass rates. Although our results show a positive impact of automatic promotion on pass rates, we find no evidence that adopting automatic progression boosts the quality of education in Brazil as the dropout rates increased after that policy as well. These outcomes suggest that the two policies by themselves do not improve the quality of education in Brazil.

Chapter 2 - This chapter examines the contemporaneous causal relationship between outdoor air pollution levels and student cognitive performance in Brazil's nationwide university entrance examinations. Our analysis relies upon a unique and not yet explored student level data set, allowing us to examine the effect of particulate matter (PM_{10}) on students' scores. To this end for our main specification we construct an individual level panel data for the two days of exams across three years and apply students fixed effects to address endogeneity concerns that arise in our analysis. In addition, we take advantage of plausibly exogenous spatial and temporal variation in PM_{10} across municipalities in the states of Rio de Janeiro and São Paulo and utilise an instrumental variable approach based

on wind direction. Our results suggest that air pollution negatively impacts the cognitive performance of students with boys more affected than girls, and less well-off exam takers at the bottom of the score distribution more affected than their more privileged counterparts. *Chapter 3* - We estimate the causal effects of acute fine particulate matter ($PM_{2.5}$) fetal exposure on Brazilian students' scores in the fifth grade using the interaction between wind direction and agricultural fires as instrument for air pollution. Our reveals find that an increase of 10% of intrauterine exposure to $PM_{2.5}$ leads to a decrease of 4.6% in students' scores in Portuguese and 2.8% Mathematics. Back-of-the-envelope calculations demonstrate that a reduction of 10% of $PM_{2.5}$ has the potential to increase later life wages by 2.3%.

Acknowledgments

I am extremely grateful to my supervisors Prof. Matthew Cole and Prof. Eric Strobl for their patient guidance. Your comments were always helpful to improve the quality of this thesis. Matt, thank you so much for taking over as my first supervisor and for asking Eric to be my second. This saved my PhD. Eric, thank you so much indeed for accepting being my second supervisor in the past nine months. I have learned a lot from you in a short period of time. Thank you for the encouragement. I am in debt with both of you.

An infinite thank you to my parents and two brothers for all the support and love. You are my backbone. Phil, thank you for being my best friend in the UK and for your support and caring. Mari and Sama thank you for decades of encouragement. Thank you my two *stinkies* Benji and Lua, my best mates ever in England, my beloved dogs! I love you all dearly.

Thank you to the editor from the Journal of the Association of Environmental and Resources Economists (JAERE) Prof. Matthew Neidell and two anonymous referees who provided excellent suggestions for the paper from my second chapter.

I would like to thank some other great people who appeared in my last year of PhD journey giving me advice, encouragement and opportunities. Dr. Davide Rigo, Dr. Livia Menezes, Dr. Liyun Zhang, Dr. Martin Koppensteiner, Prof. Olmo Silva and Prof. Tatyana Deryugina.

Thanks to the Department of Economics for the scholarship and the TEDE group for course and conference financial support. Thanks Marleen for your always swift support. Thank you my office fellows Helen, Julian, Muttamas and Raeni. Although you were not

from Economics, we passed together through the highs and lows of doing a PhD. Good luck for the four of you.

Declaration

I acknowledge that a research proposal based upon some of the work (Chapter 2) within this thesis was submitted as partial fulfillment of the University of Birmingham's Advanced Research Methods module.

Contents

Acknowledgments v

List of Figures xi

List of Tables xiv

List of Abbreviations xv

1 Introduction 1

2 Educational reforms and school quality in Brazil 5

2.1 Introduction . . . 5

2.2 Literature Review . . . 10

2.2.1 Grade retention and school dropout rates . . . 11

2.2.2 Compulsory school entrance age . . . 14

2.3 Data Description and Policy Background . . . 20

2.3.1 National tests - *Prova Brasil* . . . 20

2.3.2 School Census . . . 22

2.3.3 Descriptives . . . 25

2.3.4 Policies background . . . 30

2.4 Research Design . . . 35

2.4.1 Identification Strategy . . . 36

2.5 Empirical Findings . . . 47

2.5.1 Compulsory school entrance age . . . 47
2.5.2 Automatic Promotion . . . 54
2.6 Sensitivity Checks and Limitations . . . 56
2.6.1 Difference-in-differences with Fixed Effects . . . 56
2.6.2 Other PSM techiques . . . 60
2.6.3 Placebo Tests . . . 60
2.6.4 Limitations . . . 62
2.7 Conclusion . . . 63
Appendix . . . 64

3 The effects of air pollution on students' cognitive performance at Brazilian university entrance tests: Evidence from changes in Wind Direction 98

3.1 Introduction . . . 98
3.2 Literature on air pollution and cognitive performance . . . 102
3.3 Data and Summary Statistics . . . 108
3.3.1 Air pollution data . . . 108
3.3.2 Weather data . . . 109
3.3.3 University entrance examination scores and other ENEM data . . . 111
3.4 Empirical Strategy . . . 113
3.4.1 Panel Fixed Effects Model . . . 115
3.4.2 The instrumental variable approach . . . 116
3.5 Empirical Results . . . 121
3.5.1 Main Results . . . 121
3.5.2 Heterogeneity across sub-populations . . . 123
3.5.3 Sensitivity Analysis . . . 129
3.6 Conclusion . . . 133
Appendix . . . 134

4 Fetal exposure to fine particulate matter and students' cognitive performance: Evidence from agricultural fires in Brazil **139**

4.1 Introduction . . . 139

4.2 Data . . . 146

4.2.1 Agricultural Fires . . . 146

4.2.2 Atmospheric Conditions . . . 148

4.2.3 Air Pollution . . . 150

4.2.4 Education . . . 151

4.3 Empirical Strategy . . . 153

4.3.1 The basic OLS models . . . 153

4.3.2 The instrumental variable approach . . . 155

4.4 Results . . . 158

4.4.1 Mechanisms: Relevance of upwind fires as instrument . . . 158

4.4.2 Second Stage: Impact of air pollution on students' scores . . . 159

4.4.3 Heterogeneity . . . 160

4.4.4 School Performance and Wages . . . 163

4.5 Sensitivity Analysis . . . 164

4.6 Conclusion . . . 167

Appendix . . . 170

5 Conclusion **185**

References **208**

List of Figures

2.1 PSM balancing using only municipal schools (primary cycles) 46

A1 Simulated probability of completion sixth grade on time, *circa* 2005 74

A2 Brazil-Chile years of schooling per cohort of birth 83

A3 PSM balancing - state schools only - Method: Kernel - compulsory school entrance age 88

A4 PSM balancing - whole sample s=1 - Kernel - compulsory school entrance age 89

A5 PSM balancing -state schools s=1 - Kernel - compulsory school entrance age 92

A6 PSM balancing - municipal s=1 - Kernel - compulsory school entrance age 93

A7 PSM balancing - whole sample - Kernel - automatic promotion) 94

A8 PSM balancing - state schools - Kernel - automatic promotion 95

A9 PSM diagnostic balancing - whole sample s= 1 - Kernel - compulsory school entrance age 96

3.1 WSS, log(WSS), η^2, and PRE for all K cluster solutions (Optimal number of clusters 119

3.2 The effect of wind direction on air pollution 120

3.3 Residual PM_{10} and Residual Scores Scatterplot (student FE) 122

3.4 Cognitive impact of PM_{10} per distance from the station to the municipality centroid 131

B1 O_3 mean per day - IWD - 10km 136

B2 PM_{10} mean per day - IWD - 10km 137

B3 Rio and São Paulo municipalities centroids and pollution stations (raw data) 137

4.1 Heterogeneity in Particulate Matter's Impact on Students' Scores Across Portuguese Scores' Distribution 162

4.2 Heterogeneity in Particulate Matter's Impact on Students' Scores Across Maths Scores' Distribution 163

4.3 Cognitive (Portuguese scores) impact of $PM_{2.5}$ per distance from the fires points to the population's municipality centroid 165

4.4 Cognitive (Maths scores) impact of $PM_{2.5}$ per distance from the fires points to the population's municipality centroid 166

C1 Total of agricultural fires per municipality from 2000 to 2008 170

C2 Total of agricultural fires per municipality size from 2000 to 2008 170

C3 Total of agricultural fires per municipality from 2009 to 2017 171

C4 Total of agricultural fires per municipality size from 2009 to 2017 171

C5 Monthly average of counts of fires from 2001-2008 180

C6 Decomposition Additive Time Series of monthly average counts of fires from 2001-2008 181

C7 Monthly average of counts of fires from 2001-2017 182

C8 Share of sugarcane crop per municipality 183

C9 Share of rice, sugarcane, maize and wheat planted per municipality from 2000 to 2017 183

C10 Distribution of Fires, cropland and sugarcane in early 2000s 184

List of Tables

2.1 Frequency of schools taking *Prova Brasil* 22

2.2 Summary statistics for initial years per year 26

2.3 Summary statistics for final years per year 27

2.4 Descriptive statistics initial years - Policy: school entrance age 29

2.5 Descriptive statistics initial years - Policy: grade retention 30

2.6 Primary school nomenclature by the National Education Council 32

2.7 Whole sample - Treatment: compulsory school entrance age - (Kernel) . . 49

2.8 Only municipal schools - Treatment: compulsory school entrance age - (Kernel) 51

2.9 Only state schools - Treatment: compulsory school entrance age - (Kernel) 53

2.10 Whole sample - Treatment: Automatic promotion - (Kernel) 54

2.11 Only State schools - Treatment: Automatic promotion - (Kernel) 55

2.12 Only municipal schools - Treatment: Automatic promotion - (Kernel) . . . 55

2.13 DID-FE - Treatment: compulsory school entrance age (World Bank Method) 58

2.14 Treatment: Primary cycles DID-FE (World Bank Method) 59

2.15 PSM-DID estimations using placebo treatment: compulsory school entrance age 61

2.16 PSM-DID estimations using placebo treatment: automatic promotion . . . 62

A1 Marks scale for Portuguese exam in *Prova Brasil* 64

A2 Marks scale for Mathematics exam in *Prova Brasil* 65

A6 Matching diagnostic - treatment: compulsory school entrance age (Kernel method); cohort 1; s=0; only STATE SCHOOLS) . . . 65
A3 All the PSM techniques used . . . 71
A4 Propensity Score Matching - Difference in Differences (only states schools) 72
A5 Propensity Score Matching - Difference in Differences (only municipal schools) 73
A7 Matching diagnostic - treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; whole sample) . . . 74
A8 Matching diagnostic - treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; only state schools) . . . 78
A9 Matching diagnostic - Treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; only municipal schools) . . . 83
A10 Whole Sample - Kernel - automatic promotion . . . 90
A11 State schools - Kernel - automatic promotion . . . 91
A12 Municipal schools - Kernel - automatic promotion . . . 97

3.1 Summary statistics air pollution, weather and exams scores . . . 110
3.2 The Effects of PM_{10} on Students' Scores (Main Specification with flexible weather controls) . . . 126
3.3 Particulate Matters's Impact on Scores: Models with Instrumental Variable Wind Direction . . . 127
3.4 Heterogeneity in Particulate Matter's Impact on Students' Scores (in SD) per gender (with bins) . . . 127
3.5 Heterogeneity in Particulate Matter's Impact on students' scores (in SD) across sub-populations (bins) . . . 128
3.6 The Effects of PM_{10} on Students' Scores in SD (with co-pollutant (bins)) . 132
3.7 The Effects of PM_{10} on Students' Scores in SD (Baseline Model + 1 day lead)(bins) . . . 132
B1 Instrument for Pollution: Wind Direction – First Stage Results . . . 135

B2 The Effects of PM_{10} on Students' Scores (Specification with flexible weather controls) . . . 136
B3 The Effects of PM_{10} on Students' Scores (Linear and no weather controls) 138
B4 Number of municipalities per distance from the monitoring stations 138

4.1 Summary Statistics for main regressions sample . . . 153
4.2 Instrument for Pollution: Count of fires per Wind Direction – First Stage Results) . . . 172
4.3 The Effects of $PM_{2.5}$ on Students' Scores - OLS and IV (Main Specification with flexible weather controls) . . . 173
4.4 Heterogeneity in Particulate Matter's Impact on Students' Scores per Gender 174
4.5 The Effects of $PM_{2.5}$ on Students' Scores (using IV with different wind directions specifications) . . . 175
4.6 First Stage Results: Instruments for Pollution using non-confidence weighted fires . . . 176
4.7 Particulate Matter's Impact on Students' Scores using fires non-confidence weighted . . . 177
4.8 The Effects of $PM_{2.5}$ on Students' Scores (SD) (without and with linear weather controls) . . . 178
4.9 The Effects of $PM_{2.5}$ on Students' Scores (SD) (with co-pollutant) 179

List of Abbreviations

API	Air Pollution Index
ATE	average treatment effect
ATT	average treatment effect on the treated
CASEN	National Socioeconomic Characterisation
CCT	Conditional Cash Transfer
CEE	State Council of Education
CETESB	*Companhia Ambiental do Estado de São Paulo*
CME	Municipal Council of Education
CNE	National Education Council
DDD	Triple-Difference in-Difference
ENEM	*Exame Nacional do Ensino Médio*
FE	Fixed Effects
GDP	Gross Domestic Product

IBGE	Brazilian Institute of Geography and Statistics
Ideb	Basic Education Development Index
Inea	*Instituto Estadual do Meio Ambiente*
Inep	National Institute of Studies and Educational Research Anísio Teixeira
IV	Instrumental Variables
IWD	Inverted Weighted Distance
MDGs	Millennium Development Goals
OECD	Organisation for Economic Co-operation and Development
PISA	Program for International Student Assessment
PNAD	National Household Sample Survey
pp	percentage points
PRE	Proportional Reduction of Error
PSM-DID	Propensity Score Matching with Difference in Differences
PTA	Parallel Trend Assumption
RDD	Regression Discontinuity Design
RE	Random Effects
Saeb	Basic Education Evaluation System
SD	standard deviation

SE	standard errors
SSA	start schooling age
UN	United Nations
UNDP	United Nations Development Program
WHO	World Health Organisation
WSS	Within Sum of Squares

Chapter 1

Introduction

This chapter introduces the research topics of the thesis, which I divide into three independent studies linked by a common core issue: human capital in Brazil. It is well-established in the literature that the economic development of a country is directly related to its level of education (Romer, 1990; Mankiw et al., 1992). Therefore, studying policies aimed at improving students' performance helps policymakers to evaluate the costs and benefits of those reforms. In addition, another relevant factor that affects cognitive performance is air pollution. Thus, this thesis casts light on these issues and assesses the impact of educational policies and bad air quality on human capital in Brazil.

In the first chapter, I investigate the effects of two Brazilian educational policies on fifth and ninth graders' scores, dropout and pass rates. In the second, I study the effects of contemporaneous effects of air pollution on the students' performance at the nationwide university entrance examination in Brazil. Finally, in the third chapter, I delve into the effects of the fetal exposure to bad air quality stemming from agricultural fires in a sugarcane crop area in the State of São Paulo on the medium and long run human capital formation.

In the first piece of research, I analyse the effect of two independent educational poli-

cies on the schools' performance at the elementary school national examination *Prova Brasil*, dropout and pass rates. The main identification challenge of this chapter is the fact that both policies were not compulsory, hence raising concerns that the results could be driven by spurious correlation stemming from self-selection. To tackle this problem, I utilise propensity score matching technique with difference-in-differences to establish the causal link between each policy on pupils' performance. To do so, I build two school-level panels using rich administrative data set from the Brazilian Ministry of Education.

As per the first policy which made it optional for schools to adopt the automatic promotion from the first to the ninth grades, I found no evidence of the impact of this policy on the Portuguese and Maths' scores at the *Prova Brasil*. Moreover, I found significant and positive coefficients for pass rates, which is the aim of the practice of non-retention. However, the results indicate also a positive relationship between non-repetition in the primary school and dropout rates, which may be signalling that the policy itself does not suffice to change the alarming dropout scenario. The second policy increased the compulsory years of schooling from 8 to 9 years, changing the starting age from 7 to 6 years old. My results indicate that this policy raised the Portuguese and Maths' scores in the national examinations, but also increased the dropout rates and lowered the pass rates.

In the second chapter, I add to a vast literature on air pollution and human capital by estimating the causal effects of contemporaneous exposure to air pollution on students' performance in university entrance exams from 2015 to 2017. To identify this relationship I use students fixed effects as main specification strategy, and, as robustness checks, I utilise a wide range of control variables, such as individual characteristics, parents' level of education, household income, weather controls (temperature, humidity and wind speed), and day and municipality fixed effects. To address possible bias in our estimations stemming from classical measurement error and omitted variable bias, besides individuals fixed effects, I also perform an instrumental variable approach where I rely on the temporal and

geographical variation in wind direction to build the instruments. The main assumption is that this variation is exogenous since it does not directly affect the students' scores, except from its effect on air pollution.

Two broad findings emerge from these results. Firstly, the effect of air pollution on exam performance appears to differ for males and females, with the performance of the former more adversely affected. This finding is consistent with previous studies including Jayachandran (2009); Sanders (2012) and Ebenstein et al. (2016). Secondly, my results suggest that the effect of air pollution on exam performance differs by socio-economic background, with poorer students scoring even worse than wealthier treated exam takers. Furthermore, I run several sensitivity checks which suggest that my results are not driven by spurious correlations and seem to add plausible evidence suggesting that cognitive performance may be hindered by bad air quality.

In the final chapter, I examine the relationship between fetal exposure to air pollution from agricultural fires and cognitive performance on fifth grade exams in Brazil. To this end I merge data from the Brazilian Ministry of Education; NASA's satellite remote sensing data with active fires detection and air pollution data and atmospheric conditions gridded data from the Japanese Reanalysis archive, JRA-55. The main identification concern of this research lies on the fact that it may be possible that unobserved time-varying features may cause endogeneity in the analysis due to omission and classical measurement error inherent to this type of air pollution study. Thus, to address this problem I rely upon changes in wind direction and fires' location as predictors of air pollution, which yield credible evidence that the exposure to poor air quality during the gestational period causes a reduction in cognitive ability of students. More precisely, I argue that students living in municipalities located upwind from the fires present lower academic performance than their downwind dwellers counterparts.

My results for this chapter coincide with the previous one in this thesis as I find evidence that fetal exposure to air pollution hinders students' performance later in life. Also, when stratifying the sample per gender and scores, I find that males and weaker students seem to be more negatively affected by poor air quality. On top of that, to check the internal validity of my findings I perform a series of tests using different sets of instruments and distances between fires and municipality's population centroid, equations with co-pollutant SO_2 and falsification tests to which our results are robust and consistently suggest the harmful effect of fetal exposure to air pollution and cognitive health. In addition, I perform back-of-the-envelope calculations showing that a reduction of 10% of $PM_{2.5}$ during the third trimester of pregnancy has the potential to increase later life wages by 2.3%.

These findings shed light on the important issue of wealth accumulation mainly in developing countries with sharper levels of inequality. Suffice it to say that the less well-off, who live in areas with low environmental standards, end up suffering more from the negative impacts of fetal exposure to air pollution on their cognitive performance, and hence deepening the wage inequality in Brazil.

The next three chapters present those three studies in detail, and the last chapter concludes this thesis.

Chapter 2

Educational reforms and school quality in Brazil

2.1 Introduction

The economic development of a country is directly related to its level of education (Romer, 1990; Mankiw et al., 1992). This is recognised in Brazil whose second Millennium Development Goals (MDGs) (UN, 2015)[1] is to achieve universal primary education. As such, school attainment, quality of schooling and school dropout deserve continuous investigation from the standpoint of public policy. According to the Brazilian Institute of Geography and Statistics (IBGE) (*Instituto Brasileiro de Geografa e Estatística*), in 2008, an adult had, on average, seven years of schooling and around 15 million Brazilians aged 10 years old or more were illiterate; despite the fact that the Law 5692/1971 had increased the compulsory schooling age from 10 to 14 years old. Due to the need of understanding the mechanisms that can lead to an upgrade in the quality of education in the country, in this chapter, I investigate the effects of two educational policies that took place in 1996 and 2006 - automatic promotion and compulsory school entrance age. I study the effects of these reforms on the national examination *Prova Brasil* scores for

[1]There are eight Millennium Development Goals proposed by the United Nations (UN), and achieving universal primary education is the second target.

Portuguese and Mathematics, dropout and pass rates.

Bruns et al. (2011) reports that education has been improved in Brazil, where the average years of education has almost doubled over the last 20 years, as has the proportion of adults who have completed secondary school. Brazil has managed to meet some of the MDGs, however its students' learning performance still lags behind wealthier nations, with many high schools losing the majority of theirs pupils before graduation, presenting historically high dropout rates. In spite of the significant net educational level of the population between 7 and 14 years old in the primary school rising from 81.2 to 97.7 percent (%) from 1990 to 2012, Brazil has a long way to enhance its education quality; in particular, grade repetition and school dropout which affect the quality of schooling and are quite high[2].

Notwithstanding high expenditures in the past two decades - data from IPEA (2014) shows that, in 2014, Brazil spent 5.6% of its Gross Domestic Product (GDP) in education; i.e., two percentage points (pp) more than the investment made in 1994 (3.8% for the latter period) -, Brazil has not showed substantial advances in the returns to education, presenting levels of productivity stagnated at the levels they were in the 1980s. In comparative terms, according to Zapata et al. (2015), the country demonstrates some improvement in the Program for Program for International Student Assessment (PISA), yet quite small in comparison with other South American countries, for instance Chile. In this particular case, Brazil currently presents the level of schooling of the 1980s of the former. The National Household Sample Survey (PNAD) (*Pesquisa Nacional por Amostra de Domicílios*) IBGE (2013) and the National Socioeconomic Characterisation (CASEN) surveys Casen (2006) indicate that Brazil took almost 25 years to reach the years of schooling of Chile, i.e., in 1988, the former country achieved the mark of 10 years

[2]As shown by the United Nations Development Program (UNDP) report from 2012, one in four students who starts elementary school in Brazil leaves school before completing the last grade, which means that Brazil has the third larger dropout rate, around 24.7% among 100 nations UN (2015).

of schooling on average, a number that the latter achieved in 1965. Only by 2013, did Brazil achieve the same average in years of schooling accumulated as Chile did in 1980[3].

In addition, the students' retention impacts children and schools as the repeater feels no longer stimulated, raising the odds of dropping out of school. When one student fails a determined grade at school, not only him and his family have to bear the costs of repetition, but so do the public schools. This happens due to the fact that the institutions have to support the repeater enrolling another time at the same grade, which diverts public resources that could be used either to improve the schools' infrastructure or to admit a new student. For instance, when investing in the failed student the school wastes inputs per student as it requires more investment in school capital, such as larger classrooms, more school meals budget, desks, computers, hiring more teachers; not to mention that having older students in the same classroom may affect the younger pupils' performance in the current grade.

Some scholars think that grade retention also negatively affects the repeaters since it stigmatises them and damage their self-esteem by being forced to be among younger classmates, which usually contribute to the pupil to dropping out from school (Jimerson et al., 2000, 2002). This situation also decreases the learning achievement and leads to the postponement of the entrance in the labour market, which is another opportunity cost that the repeater will bear. Yet, part of the literature argues that grade repetition may enhance the learning achievement by exposing students who normally have lower scores to another year in contact with the same subjects, which may boost the quality of learning and education as a whole (Gomes-Neto and Hanushek, 1994; Manacorda, 2012).

The Law 9,496 of 1996 introduced automatic grade promotion as optional in from that year, with public state schools starting to adopt it in large numbers by early 2000s as

[3]See figure A2.

one of the educational policies that the government took to meet the MDGs to extend primary education for all (universal education goal) and to lower the cost of larger pupil cohorts (UNESCO, 2007). This chapter contributes to the literature on the Economics of Education by investigating how students' performance in the national examination *Prova Brasil*, dropout and pass rates are affected by two independent policies: automatic promotion and change of school entrance age from 7 to 6 years old. My study has some similarity with Koppensteiner (2014), diverging from it, though, insofar as that author looks at the automatic promotion policy in a different period of time and just one state in Brazil, Minas Gerais. Whereas my research takes advantage of the data set feature and builds upon a panel data from 2007 to 2015 (a more recent period than the previous work), comprising the whole country at the school level. Another contribution of this work to the literature relates to the investigation of students' performance in the same classroom with the same age of entrance. It means that I am not comparing the different ages within the same grade, rather looking at similar cohorts (same entrance age) at the school level. Moreover, I add to the literature by investigating the effects of a second policy concerning school entrance age on the same response variables; a topic that has been recently analysed by Rosa et al. (2019). The authors use Triple-Difference in-Difference (DDD) as main specification strategy and find the reform to have increased 5^{th} grade students' Mathematics scores by 0.1 SD and reading scores by 0.12 SD over student gains in untreated schools. [4].

Compulsory school laws have been implemented for more than a hundred years and continue to be debated. Nonetheless, the full understanding of the implications of these laws is still pending. The literature that studies grade retention norms is far from reaching a consensus whether students, schools, society and economy will be better off should they adopt automatic promotion. Moreover, the compulsory extension of years of schooling is

[4]By the date this chapter was finalised (December 2018), I was not aware of any work investigating the school entrance age policy in Brazil. It turns out that a paper was published in February 2019 analysing exactly the same topic (Rosa et al., 2019).

usually beneficial for the whole society. According to the law 11,274 of 2006, the 6 year old students should be registered at the new first grade, while the older ones should be enrolled in the former first grade[5]. Thus, in principle, older and younger children will not be in the same classroom in the first grade; which eliminates the possibility of comparison between old and young students in the same class. Therefore, my study focuses on the impact that the younger cohorts will have on the overall school performance at the school level, especially because there is no data available on the students' marks individually. The scores are released at the school, municipality and state level.

The extension of primary education to nine years of duration, with the compulsory enrolment from the six years of age, has been a goal pursued by the national education policy for many years. The nine-year primary school has extended the time of the initial years from four to five years (the final year phase consists of 4 years), to give the child a longer period for self-learning, including literacy. Besides this, the measure was taken to decrease the gap between the students from private and public schools since some private institutions had already adopted the 6 years old rule even before the law of 2006. Understanding how those two educational policies will impact the performance, the truancy, the dropout and the pass rates is the main contribution of this chapter. I quantify the benefits that dropouts forego by leaving school before the compulsory age.

To measure the impacts of both reforms on the tests scores, dropout and pass rate, I build a panel data from 2007 to 2015 at the school level, apply Propensity Score Matching with Difference in Differences (PSM-DID) technique and find evidence that automatic promotion by itself is not able to diminish dropout rates in Brazil. I find no clear evidence that non-retention is capable of tackling the quality of education in Brazil. Regarding the second policy - compulsory school entrance age - , I show evidence that Portuguese and

[5]Previously, the students were enrolled in the first grade with seven years old. I mention "new first grade" to refer to the grade to which students enrol with six years old. This distinction is important since it is not only a matter of nomenclature, but also different educational subjects taught in the new first grade.

Maths scores improved for the cohorts that were affected by the reform. For example, for the cohorts investigated in this research, my results suggest an increase in Portuguese scores between 0.35 and 0.41 SD; and between 0.14 and 0.29 of a SD in Mathematics. Notwithstanding, my findings indicate an increase in dropout rates of between 1.46 and 2.61 pp, and a fall in pass rates. These results may indicate either that the younger children are less mature and less prepared to face the first grade at the age of 6, or that the schools and staff are not well equipped to receive the younger pupils at a more demanding grade, which made more difficult to students to pass, which can be a good sign that the policy is indeed transforming, even slightly, the quality of education in public schools. Furthermore, regarding the decrease in pass rates, the *Prova Brasil* is probably easier than the internal schools Portuguese and Maths tests, leading children to fail internal exams, even though they present better marks in the two subjects in the national examination.

The rest of this chapter is structured as follows. In Section 2.2 I present the literature review. In Section 2.3 I describe the data used in this study. In Section 2.4 I present the research methodology used to assess the causal effects of the two laws on the students performance at the national examination *Prova Brasil*, dropout and pass rates. In Section 2.5 I report the main results of this empirical work. In Section 2.6 I provide sensitivity analysis. In Section 2.7 I discuss limitations and conclusions of this study.

2.2 Literature Review

The human capital approach developed by Becker (1962) is the theoretical backbone of the Economics of Education. Evaluating the returns of investment in one more year of schooling and/or the cost of grade retention in public schools to the government generates a lively literature on this economics branch. The analyses of compulsory schooling

entrance age and grade retention both theoretically and empirically casts light on the impact of these two policies on students performance, dropout and pass rates. Some other works look at the effects of these reforms on the labour market and on crime rates. Most of these studies agree that increasing the period that the children spend at school boosts their academic scores, their productivity, generating the possibility of higher salaries in the long run; or even lowering the crime rate via the incapacitating effect from taking children out of the streets as Machin et al. (2011) demonstrate.

In the next sections, I review the existing literature on compulsory schooling age and grade retention, assessing their effects on the schools performance on the national examination *Prova Brasil*, dropout and pass rates.

2.2.1 Grade retention and school dropout rates

There is no consensus in the literature on grade retention and its effects on students performance, dropout and pass rates. Some studies show that non-repetition policies lead students to be more self-confident and less stigmatised for being older than the average age within a classroom or a cohort; hence, they are less likely to dropout school (Roderick, 1994). On the other end of the spectrum, there is research showing that grade repetition enhances the quality of education, which is, in the first place, the core objective of any school reform (Gomes-Neto and Hanushek, 1994). The literature shows mixed results with sometimes positive and negative relationship between grade retention and school dropout, pass rates and exams performance. This conundrum is usually observed in developing countries, which present high retention rates and low investment in public education. Last but not least, there are also some works that point out the negative peer effect stemming from the older students (the repeaters) towards the younger (see for example Manski (1993); Lavy et al. (2011)).

Grade retention - the situation in which a student has to repeat a school year for having failed it in the previous year - from kindergarten to sixth grade is associated with a substantial increase in the probability of dropping out even after controlling for differences in background and post-repetition grades and attendance (Koppensteiner, 2014). For the United States, Roderick (1994) found that pupils who were overage at sixth grade were more likely to drop out of secondary school[6]. The author estimated that approximately one quarter of students left school, and those who remained had significant declines in attendance. These findings corroborate the thesis underpinning the present paper that automatic progression - that is, allowing students to progress from one grade to the next regardless of their academic performance during the previous year - can reduce the dropout rates ; i.e., percentage of students who leave school before the school year ends, leading, eventually, to a long run increase in the years of schooling.

Still, in a work on Chicago public schools, Jacob and Lefgren (2009) used a Regression Discontinuity Design (RDD) to assess the impact of grade retention on high school completion. The authors observed that grade failure leads to a modest increment in the odds of dropping out for older students (eight graders), whilst, for younger pupils (sixth graders), no significant effect was detected. The authors also examined the impacts of grade retention separately on those students who ultimately drop out and found no evidence that repetition changes the timing of dropout. Nonetheless, the problem of selection bias in the sample was underscored, which may have caused these counter-intuitive results.

The impact of automatic promotion on students dropping out of Uganda's primary education was investigated by Okurut (2018) using propensity score and difference in differences technique. The author found evidence that the practice managed to decrease dropout rates for those children. In examining Brazil's southern neighbour Uruguay, Manacorda (2012) worked with administrative longitudinal micro-data on secondary pupils

[6]In Brazil, secondary school is synonym of high school.

between 1996 and 1997 and, using a fuzzy RDD, the author observed that grade repetition leads to substantial dropout rate and lower educational attainment even four to five years after the grade failure first occurred. This piece of evidence builds the assumption I will make to create a dichotomous treatment variable - repetition or automatic progression - so as to evaluate whether the dropout rate will ultimately shorten due to the law of non-repetition in the early years of primary school.

Gomes-Neto and Hanushek (1994) conclude that Maths performance appears to improve more than Portuguese through grade repetition in Northeast Brazilian public schools; however, delayed promotion is quite costly - considering a full cost of 2 years of schooling. As such, the second best policy would be the mandatory promotion for the country's public school system. Whereas, the first best policy suggested is to enhance the quality of primary schools to increase the students achievement directly. Manacorda (2012) also argues that, regardless of the costs of grade retention, failing students may prevent the quality of education to decrease due to the deterrence effect of grade retention, which plays the part of an intimidating penalty that ends up inciting students to study harder to avoid failing the grade. Therefore, the threat of repetition produces positive effects not only to potential repeaters, but also to other low performance students (Koppensteiner, 2010). Consequently, new hybrid policies have appeared in the educational scheme to circumvent such a trade-off. For instance, policies that blend grade failure with remedial education, for instance summer schools, individual tutoring, support classes, etc (for more on remedial education, see (Jacob and Lefgren, 2004)). As shown in Banerjee et al. (2007) with respect to India, there were considerable pay-offs from the implementation of informal and cheap remedial education among less well-off children in that country.

All in all, this can be a plausible alternative to the grade retention - which can be an expensive policy - for other developing countries, especially for Latin American nations, whose repetition rates are higher than the Organisation for Economic Co-operation and

Development (OECD) countries, and have been repeatedly appearing at the bottom of the PISA ranking. In Brazil, there are similar policies to tackle low quality of education going on recently; for example, compensation policy - Conditional Cash Transfer (CCT)[7] such as *Bolsa Escola/Bolsa Familia* (¡for the impacts of these Brazilian CCT on education, see) Glewwe and Kassouf (2012); and *Progresa/Oportunidades* in Mexico; the P-900 in Chile, among others. *Bolsa Escola/Bolsa Familia* has shown a positive impact on the school attendance and negatively the dropout rates of the children whose families received the money.

Concerned with the quality of education after the law of non-failure, Ferrão et al. (2002) investigated the impact of that policy on students' performance from two states in Brazil: Minas Gerais and São Paulo. Using multilevel regressions, the authors found that there is no substantial loss of quality in education due to the implementation of non-repetition policies. Covering Brazil, Koppensteiner (2014) examines the effect of automatic grade progression on academic achievement in 1,993 public primary schools in the state of Minas Gerais, in Brazil, using a propensity score matching (PSM) and difference-in-differences (DID) approach. The author found a negative and significant effect of about 7 percent of a standard deviation on Math test scores; evidence in support of the disincentive effect of automatic progression. That study also found non-significant effects of the automatic progression to the dropout rates 2006 to create his treatment and control groups. Notwithstanding, the author just focused on Minas Gerais, one of the wealthiest and most educated states in the country in disregard of other regions' idiosyncrasies.

2.2.2 Compulsory school entrance age

Rosa et al. (2019)'s recent work use DDD as main specification strategy and find the reform to have increased 5^{th} grade students' Mathematics scores by 0.10 SD and read-

[7] CCT consists in giving money to the poor, conditionally to following pre-defined behaviour set by the government.

ing scores by 0.12 SD over student gains in untreated schools. Also, the literature on the topic has looked at other countries' schooling expansions using mostly student level data. Very little has been done related to the increase in years of schooling by lowering the entrance age as it happens in Brazil. For instance, in Herbst and Strawiński (2016), they use PSM-DID to investigate the reduction from 7 to 6 years old to start school in Poland, finding that older students outperform their six-year-old counterparts in writing and Maths by 0.11 ad 0.04 standard deviations, and the same performance for reading is observed.

Oreopoulos (2006) uses DID-FE techniques to assess the overall impact of compulsory schooling laws in Canada between 1920 and 1990. The author sheds light on the historical changes to compulsory schooling laws in that country and its effects on the would-be-dropouts under different circumstances than those that took place in other countries. Firstly, his main findings show that mandating education substantially increased adult income and substantially decreased the likelihood of being below the low-income cut-off unemployed, and in a manual occupation. These outcomes indicate significant gains from this sort of legislation, which seem unlikely to be offset by the costs incurred to remain in school. Secondly, the author uses changes in compulsory school laws to estimate the lifetime opportunity costs associated with dropping out of high school. Likewise, if the returns to schooling are very high, this might imply that the financial constraints or psychological costs associated with going to school must also be very substantial among dropouts. Conversely, if these explanations build upon constraints and costs that appear unlikely to hold, then they do not suffice to explain the dropout behaviour. That is, there might be other factor(s) that may be causing the truancy and the school evasion. For example, teenagers usually make their decisions based on cultural or peer influences, not to mention that youngsters tend to undervalue the returns from staying in school or even to ignore the future consequences of their decisions (Akerlof and Kranton, 2002). Therefore, these reasons and the empirical outcomes of Oreopoulos (2006) suggest that

compulsory schooling laws in Canada diminished the dropout rates.

SSA laws make parents and educators face an important trade-off. Usually, although those pupils who are the youngest in their school cohort have, on average, lower learning performance, they also have slightly higher educational attainment. Documenting evidence from a large cohort in California and Texas, Dobkin and Ferreira (2010) used RDD and observed that the school entry laws not only increased educational attainment of students who enter school early, but also lowered their academic performance while in school. Despite these results, they found no evidence that the age at which children enter school influences job market outcomes, such as wages or the likelihood of being employed, suggesting that the net effect on adult labour market aftermaths of the increased education attainment and poorer academic performance is close to zero.

Still delving into the North American continent educational laws, Peña (2017) thoroughly examined an unanticipated policy reform that shifted by four months the cut-off date for school eligibility in the state of Tlaxcal, Mexico. The author tested the effects of relative age on exam scores in grades 3 to 9, and observed that different identification strategies produce similar results: 1 year of additional age confers an advantage of roughly 0.3 standard deviations. First, using 2009 and 2012 data the author compared relative age and test scores between cohorts that started school right before and right after the reform with using DID across seasons of birth. Second, he used RDD using as the switch point the new unanticipated cutoff date for school enrolment on the 2012 data. Finally, the author instrumented relative age with assigned relative age using the 2012 data. By contrasting Tlaxcala with a neighboring state that did not change the cutoff at the same time, the study decomposed the effect of relative age into a positive effect of age at test, and a negative effect of the position in the distribution of age (for more on age at test and age at school entrance impacts on learning and attainment performance, see below) (Crawford et al., 2007, 2014). The research shows that older pupils have posi-

tive spillovers, whilst younger classmates present negative spillovers. As such, were both groups tested at the same age, the youngest would outperform their older counterparts.

Regarding South America, Berlinski et al. (2008) shed light on the effect of pre-primary education rapid expansion on pupil's ensuing school outcomes in Uruguay. The authors used a within household estimator and found small gains from pre-school attendance at early ages, which were boosted as children grew up. The work also shows that, by 15 years old, treated children have accumulated 0.8 extra years of education and these students were 27 percentage points more likely to be still on the school benches than their untreated siblings. In addition, the authors used Instrumental Variables (IV) estimates to control for self-selection of siblings into pre-school and found analogous results. Therefore, pre-primary schooling seems to be a successful and cost-effective policy to prevent early grade failure, suggesting long lasting repercussions in low income countries as it is the case of Uruguay. On the other end of the spectrum, in Chile, a one year delay appears to decrease the probability of repeating first grade and increases fourth and eighth grades test scores, according to McEwan and Shapiro (2008). The authors looked at administrative data on Chilean students that includes exact birth dates, allowing them to apply RDD to estimate the effect of delayed school enrollment on pupil results. Based on enrollment cut-offs, the study shows that a one-year postponement in school admission lowers the probability of failing the first grade by two percentage points; it also increases fourth and eighth grade test scores by more than 0.3 standard deviations, with larger effects for boys.

Goodman and Sianesi (2005) investigated the effects of taking early education - before the compulsory starting age of 5 - and of pre-school on a cohort of British children born in 1958[8]. Controlling for child, parental, family and neighbourhood characteristics, they found some positive and long-lasting effects from early education. Particularly, pre-compulsory education appears to significantly upgrade cognitive tests at age 7 at a

[8]Excluded Scotland.

small magnitude, remaining significant up to age 16. Whereas, attendance of pre-school - nursery or playgroup yields a positive though short-lived impact on test scores. Not only better scores are positive outcomes from compulsory schooling age reform, but also socialisation effects are found positive and short-lasting.

An important aspect of start schooling age (SSA) policies is covered by Crawford et al. (2007), who argue that it is the age at the tests effect that matters the most, instead of the SSA. Implementing RDD, the authors examine the impact of month of birth on national achievement test scores, finding small negative effect of SSA at age of 7, but this impact is dwarfed by the age-at-exam effect. Furthermore, they found that the SSA or length-of-schooling effect has disappeared by age 14. The authors complemented that work with another study Crawford et al. (2014) and using RDD again they estimated an upper bound of the effect of age at test. They found that the difference in age at which cognitive skills are tested accounts for the majority of the difference in the outcomes above.

Looking at Norway's SSA reform, Black et al. (2011) found small and negative effect of starting school older, although much larger positive effects of age at test. Besides that, starting older leads to lower earning until the age 30. The authors also found little impact of SSA on education attainment, but boys who start older are less likely to have poor mental health at the age 18; as well as negative effect of starting older and teenage pregnancy. Their analysis was build up at the students level, with a very defined cut-off date and change in trajectory impacts, allowing the use of a RDD to delve into the aftermaths of that SSA reform. Notwithstanding these findings, another study on Norwegian compulsory kindergarten effects stemming from an SSA changing the age from 7 to 6 year old did not find any effects on later school performance (Drange et al., 2012).

Studying Germany SSA, Fertig and Kluve (2005) used PSM-DID to capture the effect of age at school entry on educational attainment. The author found a negative rela-

tionship between the age the children start school and their learning outcomes, both in terms of school degree and probability of grade repetition and higher likelihood of dropout school. More recently, Betthäuser (2017) assessed the effects of compulsory schooling law change from 8 to 9 years in Germany by using DID and found evidence that the reform narrowed the gap in educational performance between different social origin groups. As a consequence, the reform also produced impacts in the long run as the author observed a reduction in the inequality in the labour market, leading to a higher social mobility in the country.

Interested in the effects China Compulsory Education Laws from 1986, Zhang et al. (2017) estimate the effect of school starting age on academic achievement on junior high school students. The authors apply RDD to estimate the treatment effects on enrolling at school at a cut-off point, by comparing the outcomes near the threshold, which is the child's age on September 1st in the school year that she entered primary school. The authors demonstrated that one year delay is associated with a decrease in with a 0.303 standard deviations of cognitive scores, oppositely to many Western countries results. However, this negative effect is caused by human capital accumulation prior to primary school entry. Firstly, the authors demonstrated a larger negative effect for children from rural schools and a small effect for those from urban schools. Secondly, the SSA is negatively related with the length of pre-school attendance. The author conclusions suggest that wealthier families may be able to better offset any negative impacts from starting school at a younger age by investing more in children before they enter primary school. Moreover, due to the lack of pre-schools in China, poor families are more prone to keep their children at home before they can start primary school. While, the children from richer households live in a more stabilised family environment and attend higher-quality pre-school. Hence, along with the preference for early school enrolment, the entrance cut-off aggravates the socio-economic gap in school performance.

2.3 Data Description and Policy Background

For this chapter, I build a biennial panel using administrative data at the school level from 2007 to 2015 with 55,567 primary public schools located in Brazil to assess how the implementation of two educational policies affected the test scores of municipal and state schools in the *Prova Brasil* examinations; the pass and the dropout rates. I merge different data sets from various official websites to collect the dependent, the main explanatory variable - the treatment - and the other control variables as described in the next subsections.

2.3.1 National tests - *Prova Brasil*

The scope of the *Prova Brasil* is to evaluate the quality of education in public schools in Brazil, comprising Portuguese and Mathematics tests and a socioeconomic questionnaire, taking place every two years (in the odd years). To minimise costs and make the evaluation easier for the Ministry of Education, not all the public schools in the country take part in this examination. Besides this, the tests are not compulsory, i.e, there are schools and students that do not participate. to take part in the tests, schools must have at least 20 students enrolled in the assessed grades. The pupils should pass a Portuguese and a Mathematics test, at the 5^{th} and 9^{th} grades, and the scores are given at the school level, per states and per municipalities.[9]

The *Prova Brasil* is created by teachers and researchers from the whole country, who elaborate questions following specific instructions from the Ministry of Education. These guidelines present the set of skills that should be assessed by the tests per grade and disciplines. Unlike the exams applied by the teachers in the schools, the methodology adopted by the *Prova Brasil* addresses the evaluation of the schools' and the educational system's

[9]Brazil is a federation composed of the union of the 26 states, the Federal District, and the 5,570 municipalities. The latter are administrative divisions of the Brazilian states which have, on average, 214 municipalities (Wikipedia, 2007).

quality rather than the students individually. Likewise, the guidelines and the Basic Education Evaluation System (Saeb)[10] marking scale (see Tables A1 and A2) are the same for all the editions of the exams allowing the comparison between results from different years.

Overall, there are 21 different types of questions booklets. Considering that the classes have approximately 21 students on average, pupils end up not answering the same questions, which lessens the chances of students be prepared by their teachers for the tests. The questions are of multiple choice, with 4 or 5 items each, and only one correct. The scores are divided by skills levels; ranging from 0 to 425 marks (elementary school goes up to 400) for Portuguese, and from 125 to 475 (elementary school up to 425) for Mathematics. Table 2.2 summarises the records of treated and untreated schools in the *Prova Brasil* Mathematics and Portuguese tests.

From 2007 to 2015, 55,567 public schools participated of the *Prova Brasil* for fifth graders per year; while only 40,334 schools took the tests for the ninth year. This is due to the fact that there are fewer schools for the last years of primary school than for the first years. Moreover, considering the increase in dropout rates after the 5^{th} grade, fewer students take part in the national examination at the ninth grade. Still, the same schools that took the exam for the the 5^{th} grade do not necessarily take it for the the 9^{th} grade in the same year. Hence, if I merge the data set with the *Prova Brasil* results from initial years (the 5^{th} grade) with the final years (9^{th} grade) in the same panel, I end up with only 26,511 schools that took both exams in the same year as demonstrated in the Table 2.1 below.

In light of that, from the *Prova Brasil* data set, I take the Portuguese and Maths results and pass rates to use as dependent variables in this study. Thus, to keep as much

[10]Saeb is composed of a set of large-scale external evaluations that allow the National Institute of Studies and Educational Research Anísio Teixeira (Inep) (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*) to make a diagnosis of Brazilian basic education and some factors that may interfere with the student's performance, providing an indication of the quality of the education offered.

Table 2.1: Frequency of schools taking *Prova Brasil*

	5th grade	9th grade	both exams
Year		Frequency	
2007	55,567	40,334	26,511
2009	55,567	40,334	26,511
2011	55,567	40,334	26,511
2013	55,567	40,334	26,511
2015	55,567	40,334	26,511

Note: Table elaborated with information from Basic Education Development Index (Ideb) (*Indice de Desenvolvimento da Educacao Basica*) microdata. Note that there are more schools registered to take the fifth year exams than the fifth year, and that the number of schools that took both exams in the same year drops by around half.

data as possible, I create three different panels: (i) one with just initial years (4^{th} or 5^{th} grades), which I merge with the data set from school census - which bring information on schools, teachers and students - just for first years; (ii) a second one with the same information, but for final years (8^{th} or 9^{th} grades); and (iii) a third panel with less data as it includes only the schools that took the *Prova Brasil* at initial and final years. I also use the dropout rates as response variable, which I collected from the school census and merged at the school level with these panels.

2.3.2 School Census

The Inep, an agency linked to the Ministry of Education, carries out the School Census, an annual nationwide statistical-educational survey. The census inquires into basic education information, such as number of enrolments; pass and dropout rates; students income; pupils' parents' (or legal tutors') characteristics; schools' infrastructure and teaching functions. The data are provided by the public and private schools themselves, state and municipal education systems. This unique and rich data set provides the dropout rates - herein used as another dependent variable - as well as the control variables adopted to conduct the propensity score matching and the econometric models. For example, par-

ents' educational level, gender, age in the date of the examinations, children work and household details.

Treatments: Compulsory school entrance age and automatic promotion

The school census provides the data for the two treatment variables, each of them accounting for the automatic promotion and for the change in the compulsory school entrance age. The survey asks the schools if they adopt the non-retention policy in the year it is being conducted; and if the school opted to enrol children aged 6 years old in the first grade. The non-repetition policy is optional; that is, the educational systems municipal or state decide whether the schools under their control will opt for automatic promotion, and the duration of this decision. As a result of the decision being made by the system rather than by the schools themselves, the concern regarding the possibility of endogeneity stemming from that option can be relaxed here.

As per the change in the compulsory school starting age, the government allowed a period of transition and adaptation for the schools to welcome the younger pupils. Thus, from 2006 to 2010, the state and municipal schools could choose if they would receive the younger students in the first grade; and, from 2010 onwards, the new starting age of six years old started to be compulsory for all school in the country. The optional phase is of the utmost importance for the public schools since it enabled them to train the teachers to cope with younger pupils in the first grade; to adapt the classrooms, furniture and course materials suitable for the younger students cohorts.

Variables for matching and control

I use a number of variables to carry out the propensity score matching and to control schools' idiosyncrasies for as many observables as possible and to try and minimise the

possibility of omitted variable bias. The following variables were also included in the models to identify the common pre-trends trajectory, which is a crucial assumption of the difference-in-difference technique: students' characteristics per school: gender; age; pre-school; race; work; ever failed; live with parents; mother schooling; parents' interest in children education (see Appendix for more detailed variables description).

To transform the individual-level variables into proportions per school, I utilise student level data, by collecting the data from the socio-economic questionnaires from *Prova Brasil*; and from the School Census students data set. I collapsed by proportion each item of the questionnaires and transformed them into a pupil level panel data so as to merge with the school level panel data. Hence, I created the proportions of each characteristic of interest per school, accordingly. Despite the existence of such information at the student level, there is no information at the same level for the marks, pass and dropout rates, which is the reason why I could not merge and analyse this entirely at the student level.

After merging the data sets, I remove from the sample all the inconsistent observations and missing or incomplete data. Furthermore, I drop the federal schools - schools that receive funding from the federal government, hence wealthier than state and municipal funded schools - from the samples as they traditionally display higher scores than the other public schools and would differ from the rest of the sample. I also dropped the high schools, the special education, the distance learning, and the adults education so that the remainder observations are only the state and municipal schools, with presential 5^{th} and 9^{th} grades. Private schools are not included in my investigation because they were not obliged to pass the national examinations - *Prova Brasil*.

2.3.3 Descriptives

The *Prova Brasil* and the Ideb data sets bring the examination results separating the schools into two groups: initial years, from first to fifth grades; and final year, from sixth to ninth grades. Following this structure, I split the main summary statistics into Tables 2.2 and 2.3. Table 2.2 demonstrates that the pass rates increased over the period from 2007 to 2013, though, from 2013 to 2015, there was a drop of approximately 13.71 pp. The same pattern is seen for Portuguese and Maths scores: the marks increased in the same interval of time and decreased from 2013 to 2015 (Portuguese around 3.89 pp lower; and Maths, 5.6 pp). Regarding the dropout rates, similar movement occurred: there was a negative trend between 2007 and 2013, with a change of slope to an increase of the dropout rates from 2013 to 2015 of roughly 10.69 pp.

Note that the number of schools change over time and across response variables. This is due to the fact that not all the same schools take exams every year and not all of them have their scores revealed. Besides this, the dropout and pass rates come from a data set different from the *Prova Brasil*'s, which means that some observations were dropped when merged as sometimes the schools have their scores available for one year in one data set but do not have for pass and dropout rates. In addition, the number of schools that take the examinations in the fifth grade is larger than the ninth grade. All the data sets I used to match show fewer ninth grade schools than fifth grade ones, certainly because there are fewer schools for the higher grade as there are fewer students that achieve the ninth year (mainly because the dropout rates increase after the fifth grade). Another explanation, for example, is that one school may have taken the fifth grade exam in a given year, though not the final exam in that same year.

When it comes to final years, pass rates have increased across the whole period and dropout rates have dropped from 2007-2013, but slightly increased by 7.23 pp from 2013-2015. The performance in Maths and Portuguese have oscillated throughout the years as

Table 2.2: Summary statistics for initial years per year

Initial yrs		2007	2009	2011	2013	2015
Pass	Mean	86.37	87.58	90.17	91.55	90.11
	Sd	10.67	10.06	8.66	7.84	8.21
	Min	0.9	0	0	29.6	19
	Max	100	100	100	100	100
	Obs	35,855	42,109	39,198	37,601	22,698
Dropout	Mean	2.95	2.35	1.67	1.31	1.45
	Sd	4.88	3.99	3.03	2.38	2.44
	Min	0	0	0	0	0
	Max	81.8	100	100	31.8	39
	Obs	35,905	42,187	39,447	30,839	22,698
Port	Mean	169.73	192.33	193.71	197.94	190.25
	Sd	18.77	21.5	22.13	25.96	23.85
	Min	103.67	121.64	106.06	104.95	106.06
	Max	261.57	243.33	243.98	315.74	314.75
	Obs	28,650	34,574	32,622	30,839	17,311
Maths	Mean	187.17	212.56	211.8	214.58	202.54
	Sd	20.56	28.15	29.31	35.8	23.17
	Min	118.17	104.97	104.97	10.59	133.86
	Max	320.88	342.32	339.22	342.32	342.32
	Obs	28,650	34,572	32,622	30,847	17,311

Note: This Table shows contains the summary statistics for the dependent variables Pass and Dropout rates, Portuguese and Maths' scores. Initial years comprise from first to fifth grades.

shown in Table 2.3.

Table 2.3: Summary statistics for final years per year

Final yrs		2007	2009	2011	2013	2015
Pass	mean	78.49	79.83	82.34	83.46	83.86
	sd	12.57	12.23	11.99	11.35	11.37
	min	12.7	0	0	0	0
	max	100	100	100	49.3	100
	obs	15,561	17,168	16,040	14,415	22,854
Dropout	mean	6.57	5.4	4.38	3.46	3.71
	sd	8.02	6.67	5.56	4.25	6.1
	min	0	0	0	0	0
	max	70.1	100	100	49.3	100
	obs	15,617	17,213	16.098	14,415	11,685
Port	mean	225.09	243.4	236.95	236.1	235.79
	sd	18.44	7.65	15.69	14.87	15.93
	min	139.6	232.2	157.4	156.4	156.4
	max	313.9	249.2	279.9	281.1	281.7
	obs	15,561	17,166	15,954	14,416	20,974
Maths	mean	237.16	247.96	251.94	247.36	247.3
	sd	18.44	8.82	19.08	15.8	17.46
	min	139.6	239.84	180.04	189.9	189.9
	max	313.9	260	318.3	309	309.5
	obs	15,561	17,166	15,954	14,416	20,974

Note: This Table shows contains the summary statistics for the dependent variables Pass and Dropout rates, Portuguese and Maths' scores. Final years comprise from sixth to ninth grades.

Additionally, Table 2.4 brings the behaviour of the response variables separated by treatment and control groups (policy: school entrance age) and per year between 2007 and 2015. The policy is from 2006, but there is no data available on the response variables for that year; therefore, I start my panel from 2007. I consider the first cohort to be affected by the new school entrance age the one which starts school at the age 6 in 2007 and take the fifth grade national exams in 2011. Hence, the effects of the policy can be noticed in 2011 for the first cohort doing the fifth year tests, and in 2015 for the same cohort doing the exams at the ninth grade. Consequently, one cannot observe any impact of that policy before the year 2011 as all students that took the exams in 2007 and 2009 are not affected by the lower entrance age, i.e., they started school at the age of 7. Also,

Table 2.5 brings the summary statistics for the treated and control groups with respect to the grade retention policy; where, retention refers to the control group and non-retention, for the treated schools. The figures and tables indicate a change in the means after the implementation of the policy.

Table 2.4: Descriptive statistics initial years - Policy: school entrance age

Year		2007		2009		2011		2013		2015	
Dep.Var.	Yrs.sch	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
	8 years	84.54	12.49	82.99	12.74	83.56	12.96	83.9	11.37	86.87	9.69
Pass	Obs	12,197		5,920		632		89		403	
	9 years	87.32	9.46	88.33	9.34	90.27	8.52	91.57	7.82	90.16	8.17
	Obs	23,658		36,189		38,566		37,512		22,295	
	8 years	3.46	5.72	3.89	5.45	3.97	7.07	2.92	4.3	2.44	3.54
Dropout	Obs	12,217		5,937		644		89		403	
	9 years	2.69	4.35	2.1	3.63	1.63	2.91	1.31	2.37	1.43	2.41
	Obs	23,688		36,250		38,803		37,512		22,295	
	8 years	168.84	18.3	184.18	19.56	182.93	23.32	175.65	18.71	180.36	23.67
Port	Obs	11,325		5,793		622		88		386	
	9 years	170.31	19.05	193.97	21.49	193.92	22.05	198	25.95	190.48	23.81
	Obs	17,325		28,781		32,000		30,751		16,925	
	8 years	186.2	20.14	204.05	22.89	197.46	28.19	197.4	19.72	193.77	22.26
Maths	Obs	11,325		5,792		622		88		386	
	9 years	187,81	19.05	214.28	28.79	212.08	29.27	214.63	35.82	202.74	23.15
	Obs	17,325		28,780		32,000		30,759		16,925	

Notes: Pass and dropout rates range from 0 to 100 in percentage. Port stands for Portuguese results and range from 0 to 400 marks. Maths results range from 125 to 425 marks. Yrs.sch stands for the duration of schooling: 8 years non-treated and 9 years treated.

Table 2.5: Descriptive statistics initial years - Policy: grade retention

	2009		2011		2013		2015	
Treat/control	mean	sd	mean	sd	mean	sd	mean	sd
retention	85.1	10.26	88.23	8.87	90.08	7.9	89.84	8.13
Obs	26,166		23,974		22,995		16,254	
non-retention	91.64	8.24	93.21	7.35	93.86	7.15	90.76	80.36
Obs	15,943		15,224		14,606		6,444	
retention	2.89	4.38	2.05	3.37	1.56	2.56	1.47	2.48
Obs	26,226		24,180		22,995		16,254	
non-retention	1.46	10.26	1.08	2.3	0.92	2	1.38	2.35
Obs	15,961		15,267		14,606		6,444	
retention	185.16	20.35	188.02	21.91	191.5	25.23	188.17	23.83
Obs	19,574			18,287	17,354		11,620	
non-retention	201.7	19.23	200.98	20.2	206.23	24.5	194.5	23.33
Obs	15,000		14,335		13,485		5,691	
retention	204.37	24.39	204.65	27.64	209.88	10.59	200.43	23.06
Obs	19,575		18,829		17,355		11,620	
non-retention	223.25	29.12	220.92	28.85	220.63	39.98	206.85	22.81
Obs	14,997		14,333		13,492		5,691	

Notes: Retention stands for the schools that did not opt for automatic promotion; and non-retention for school that did adopt it.

2.3.4 Policies background

The law 11,274/2006 has defined an optional period from 2006 until 2010, when schools could opt to follow it or not. The optional period was set to allow the schools and the public educational system to meet all new entrants' demand, adequacy of physical and pedagogical material, number of teachers and administrative support, appropriate training and career staff plan. Most public schools need to adapt not only their infrastructure, but also their staff to receive the younger pupils in a grade that used to welcome seven-year-old children. This includes the re-organisation of elementary school, for first and nine years.

Compulsory school entrance age

The law 11,274/2006 sets out an optional period from 2006 until 2010, when schools could decide whether they would enrol six-year-old children in the 1^{st} grade. The new cut-off date for children to start the primary school is 6 years old completed or to be completed by the start of the school year, which, in Brazil, is in March. In principle, parents can choose which public school they will take their child to according to their preference. For instance, some parents - the more conscious of the importance of education - apply for schools that present high Ideb scores, [11] regardless of how far it is from their houses. It turns out that the schools have some criteria for the order of priority to accept the child. In addition to the proximity between the student's residence and the institution, having siblings in the school where the application is made is also a factor that the schools take into consideration to make the admission decision (for a thorough discussion on distance between school and home, and the endogeneity of the school choice decision by parents, see Card (1999, 2001)). This is due to the scarcity of places for all children in the primary school in the public system in Brazil, which causes enormous queues outside the schools during the enrolment period. Some parents have even to spend the night before the first day of registrations in the school street so as to try to guarantee a place for their children, although queuing does not assure that they will find a place in that particular institution.

When it comes to school entrance policy, usually it implies having children who are at different ages in the same class or grade, which may entail some cognitive consequences in the long run (Crawford et al., 2010; Dobkin and Ferreira, 2010; Black et al., 2011;

[11] Ideb is the main quality indicator of basic schooling in Brazil. Ranging from 0 to 10, it synthesises two concepts: the schools' pass rates and the learning level on Mathematics and Portuguese gauged by the *Prova Brasil*, the national examination for students from primary to high school, including adult education. The Ideb calculation is represented as follows:

$$\text{Ideb}_{ji} = N_{ij} * P_{ij}; 0 \leq N_j \leq 10; 0_j \leq 1; \tag{2.1}$$

where, $i = Ideb$ and School Census year; N_{ij} is equal to the average of Portuguese and Maths normalised average for a indicator between.

Peña, 2017). To mitigate this concern, the law 11,274/2006 establishes a cut-off date to enrol at the primary school, which is 6 years old completed or to be completed by the start of the school year. Hence, the new classes that will receive 6-year-old pupils do not enrol children with higher ages, except the students who repeated the first year, i.e., coming from a previous cohort. Furthermore, those who started their 8-year primary school - the previous cohorts - will not benefit from the 9-year elementary school. Likewise, six-year-olds who can read and write cannot be directly enrolled at the second year of the nine-year primary school since increasing the compulsory years of schooling objectives to extend the time and quality of the teaching-learning, instead of anticipating its conclusion. Additionally, both eight-year and nine-year curricula should coexist, during the former's process of extinction and the latter's implementation/transition phase.

The new nomenclature suggested by the National Education Council (CNE) (*Conselho Nacional de Educação)* for child education is displayed in the Table 2.6 . For more details on the Brazilian political-legal education system reforms, please see the Appendix.

Table 2.6: Primary school nomenclature by the National Education Council

School level	**Suitable starting age**	**Duration**
Child education	Until 5 years old	
Nursery	Until 3 years old	
Kindergarten	4 and 5 years old	
Primary School		9 years
Initial years	From 6 to 10 years old	5 years
Final years	From 11 to 14 years old	4 years

Note: Table elaborated with data from Basic Education Secretariat - Ministry of Education of Brazil SEB/MEC (2006).

The Resolution CNE/CEB No.3/2005 spells out that children who are up to 5 years old in the beginning of the school year must be enrolled at the kindergarten. Furthermore, the child education stage will be able to take care of children under 6 years old in two situations: (i) until the education system extends the Primary School to nine years of

duration respecting the deadline (in 2010) set by the law; and (ii) to reach all children who turn 6 after the date set for entry into schooling.

Regarding the municipalities, initially, they must consider whether they are linked to the state educational system rules or if they had their own system. In the former case, the municipality must comply with the State Council of Education (CEE) (*Conselho Estadual de Educação*) and submit its reform proposal for the Elementary School to analysis and approval of the CEE. The municipality with its own education system must just observe the norms already updated by its respective Municipal Council of Education (CME) (*Conselho Municipal de Educação*).

Automatic promotion

The Law 9,496 of 1996 introduced automatic grade promotion as optional in that year, with public state schools starting to adopt it in large numbers by the early 2000s.[12] Each educational system decided whether it was convenient to establish it in its premises. The organization of the public education system in Brazil is based on boosting the quality of teaching and learning in all schools, equality and equity for all students, access and permanence in all basic education. This expected improvement in the quality of educational services depends on the joint effort of its professionals for the implementation of public policies of education and articulation between the State Department of Education, Teaching Offices and schools of the state public education network. Each state or municipal network is entitled to opt for non-retention or not, and the schools linked to these systems should follow their guidelines accordingly. For instance, the State of São Paulo has adopted in its public education system the organization of elementary education in cycles, obligatory for the state public network. The Resolution SE 53 of October 2^{nd}, 2014 provides orientation for the re-organisation of the Elementary School in Continuing Pro-

[12]This section is based on information gathered by electronic contact to the Secretariat of Education of São Paulo, which explained the dynamic of the automatic promotion in that state, which I am roughly generalising here to all other states that adopt it.

gression Regime and on the Mechanisms of school support to the Elementary and Middle School students of the state schools, creating 3 learning phases, with a duration of 3 years each. The division of teaching in three cycles of learning aims to provide pedagogical conditions for children and adolescents to obtain more opportunities to be effectively served in their needs, enabling them to adapt learning times to their individual characteristics. The learning cycles, understood as time spaces interdependent and articulated among themselves, are defined throughout the nine years of Elementary Education as follows: i) Literacy Cycle, from 1^{st} to 3^{rd} year; ii) Intermediate Cycle, from the 4^{th} to the 6^{th} year; and iii) Final Cycle, from the 7^{th} to the 9^{th} year.

Failure in primary education has increased over the years, hence, to overcome the so-called "pedagogy of repetition" and exclusion, and to establish a pedagogy of human promotion and inclusion, the government decided to make it optional for the educational systems throughout the country to opt for continuous progression in elementary education, acknowledging its contribution to enhance the quality of teaching and learning. This strategy is an effort to fulfill the second MDG of universalising the basic education, to guarantee the access and permanence of the children at school and to regularise the flow of students in consonance with the relation age-grades.By adopting the system of continuous progression, the policy maker recognizes that the cycle exceeds the learning time in a single school year, stimulating less truancy and considering the students' idiosyncrasies. As a result, the scope of the students evaluations become an indicator of monitoring the learning progress, instead of being a mere instrument of pupils approval or disapproval.

As mentioned above, the public schools do not make the decision to adopt or not continuous progression (or automatic promotion); the choice is made by the municipal or state network to which the school is linked. As such, as my study is at the school level, rather than the municipal or state systems, I can avoid endogeneity concerns due to this specific issue.

2.4 Research Design

I focus on the schools' performance in the national examination *Prova Brasil*, pass and dropout rates, before and after changes in compulsory school entrance age[13] and automatic promotion laws. First, I examine whether schools who opted to enrol children aged six years old - instead of 7 years old, as it used to be before the new law 11,274/2006 - in the 1^{st} grade during the optional phase (2006-2010) of the reduction of school entrance age policy presented better results in Maths and Portuguese tests, higher pass rates and lower dropout rates. Second, I investigate if the schools who adopted automatic promotion in, at least, one year from 2011 to 2015 showed increased results regarding these same dependent variables.

One may question the possibility of self-selection bias stemming from the fact that students are not obliged to participate in the exam, although the schools are oriented to encourage and ensure that everyone is present on the day of assessment, including the ones with disabilities. One could also argue that the more responsible and hard working pupils would be more likely to be present in the day of the exams, or that the students may be given an incentive by the head of the school in terms of marks in their own school examinations if they take the *Prova Brasil* tests. Notwithstanding these valid concerns, it does not seem to be a serious problem since the national examinations are constructed to assess the schools, not the students themselves. Regarding the potential existence of a reward, in case it really happens, it would be offered to all students, thus the endogeneity problem here could be mitigated under this aspect.

Moreover, to avoid upward bias brought about by the self-selection of parents choosing where to enrol their children, I control for the mother's level of education and for the

[13] I analyse its first optional phase as a quasi-experiment.

proportion of parents that are interested in their children education.[14] Hence, I created these variables from the socio-economic questionnaire the pupils fill in during the *Prova Brasil* tests. My assumption is that the more educated she is and/or the more involved in the child education the parents are, the larger the chances of their decision be guided by their preference for better education.

2.4.1 Identification Strategy

Following the literature on education reform policies (Marques da Costa, 2018; Machin and Wilson, 2005; Eyles and Machin, 2015; Koppensteiner, 2014; Fertig and Kluve, 2005; Herbst and Strawiński, 2016; Carvalho, 2014; Okurut, 2018), to avoid endogeneity and biased estimates stemming from schools self-selecting to adopt the policies, I use Propensity Score Matching with Difference-in-Difference (PSM-DID) technique (Lechner, 2002; Leuven and Sianesi, 2003). The matching routines are able to remove that particular bias (to some extent) given that it creates a valid control group based on the differences between the schools that adopted the policies and the schools that did not.[15]

My study diverges from the aforementioned as I analyse Brazilian public schools, which has quite different institutions and school systems from England and Poland, countries considered in those empirical works. Additionally, I consider both reforms in my models; controlling for automatic promotion in the regressions whose main treatment is the school age reform. This study's purpose is to estimate the average treatment effect on the treated (ATT) for each reform so as to appraise the difference in public schools' performance at the *Prova Brasil* examinations, dropout and pass rates between schools which have adopted one or both of the treatments and equivalent schools which have not opted

[14] In the appendix, I list 4 questions that the students answer regarding their parents or legal tutors interest in their education: parents' incentive to attend the classes; parents' incentive to do the homework; parents' incentive to read and parents' incentive to study.

[15] In the methodology section, I show the limitations of the PSM technique concerning the unobserved characteristics that change over time.

for the treatments. Furthermore, I am able to evaluate not only the static results of the policies implementation, but also their dynamic effects by looking at two periods (t and $t+1$) and two cohorts. The latter analysis can just be done for the school admission age policy; the automatic promotion treatment does not provide me with enough data to assess more than one cohort.

The causal effect of lower school starting age and automatic promotion on the schools' performance at time $t+s$ can be identified as:

$$Y^1_{i(t+s)} - Y^0_{i(t+s)} \tag{2.2}$$

where I consider y_{it} as school i's scores in the *Prova Brasil*, dropout or pass rates at time t; and $Y_{i(t+s)}$ is the scores in the *Prova Brasil*, dropout or pass rates s period(s) later, with $s \geq 0$. The superscripts stand for the school decision: 1 if a school adopts the policy at t and 0 otherwise. Consequently, $Y^0_{i(t+s)}$ is the school i performance at period $t+s$ had it not opted for the new policies since period t.

Assignment to treatment

Although the ideal set to assess the causal relationship between the treatment and the response variables - the treatment effect - are randomised trials, most empirical studies rely on quasi-experiments to create the counterfactual group. My study sheds light on the optional phase of the school entrance age law, therefore I use matching methods to emulate a randomised experiment and to reduce the self-selection bias coming from the decision of the schools to adopt the new entrance age before the compulsory phase of the new policy. Since I cannot observe what would have happened had the treated schools not opted to lower the school age during 2006-2010, I have to create a relevant counterfactual to assess the causal effect from the difference between the outcomes of schools not exposed

to the treatment (counterfactual) and the schools actually exposed to the treatment (the factual). In this research, the outcomes of interest are (i) educational attainment based on the *Prova Brasil* scores; (ii) dropout rate; and (iii) pass rate. The treatment is school entrance age at 6 years old rather than at 7 years old; and automatic promotion (non-retention).

The main challenge to define the counterfactual is due to the fact that only one of the two aftermaths of Equation 2.2 is actually observed. For instance, if the school i decides to comply with the policy, its outcome variable will be y^1; the y^0 (the counterfactual result) cannot happen. As a result, I am not able to estimate the direct treatment effect for the treated schools. Then, I use the population average treatment effect (ATE) as follows:

$$ATE = E[Y^1_{i(t+s)} - Y^0_{i(t+s)}] \tag{2.3}$$

i.e. the difference in the expected responses of treated and non-treated.

Propensity Score Matching

I use PSM approach to select a valid control group from the untreated schools and to keep in my sample just the control schools with similar observed characteristics from the pre-policy period to the treatment group's (Becker et al., 2002); hence, minimizing the omission bias that my estimations would incur had I not matched on the observables with the PSM (Rosenbaum and Rubin, 1985; Heckman et al., 1998). After matching, I drop the schools from the control group that are not in the support group and estimate the ATT just with the schools from the control group and the treated schools. Thus, the ultimate matching goal is to associate each treated school with a school that has never adopted the policy given a set of observable characteristics.

I capture the difference between schools' performance after a school opts for the educational policy, by creating a dummy variable $TREAT_{it}$ (standing for beginning of the policy), which equals 1 if the school starts to adopt the policy at period t, and 0 otherwise. The average effect on treated schools' (the schools that lowered the entrance age to 6 years old or the schools that adopted non-retention) marks in the national examinations and on their dropout and pass rates is the ATT, which is expressed as follows:

$$\begin{aligned} ATT &= E[Y^1_{i(t+s)} - Y^0_{i(t+s)} \big| TREAT_{it} = 1] \\ &= E[Y^1_{i(t+s)} \big| TREAT_{it} = 1] - E[Y^0_{i(t+s)} \big| TREAT_{it} = 1] \end{aligned} \tag{2.4}$$

Similarly, the counterfactual is denoted by $E[Y^0_{i(t+s)} \big| TREAT_{it} = 1]$, but it is not observed, as explained above. Therefore, the average of the dependent variable regarding the control group (the schools that did not adopted the lower school entrance age or the automatic promotion) is a good substitute for the unobserved counterfactual (Heckman et al., 1998). It follows that Equation 2.4 is then:

$$ATT = E[Y^1_{i(t+s)} \big| TREAT_{it} = 1] - E[Y^0_{i(t+s)} \big| TREAT_{it} = 0] \tag{2.5}$$

Firstly, I conduct the PSM, which estimates the probability that a school adopts the policy in time t, based on their observable characteristics. To estimate the likelihood I use probit model as follows:

$$\begin{aligned} Pr\big(TREAT_{it} = 1\big) = \psi\big(&gender_{i(t-1)}, age_{i(t-1)}, presch_{i(t-1)}, race_{i(t-1)}, work_{i(t-1)}, failed_{i(t-1)}, \\ &liveparents_{i(t-1)}, mothsch_{i(t-1)}, parentsinterest_{i(t-1)}, pcs_{i(t-1)}, internet_{i(t-1)}, labs_{i(t-1)}, \\ &teachersroom_{i(t-1)}, teacherSdiploma_{i(t-1)}, D_t\big) \end{aligned} \tag{2.6}$$

where Pr denotes the predicted probability of school i adopting the policy at time t; and $\phi(.)$ is the normal cumulative distribution function; and D_t is a time dummy to capture time effects. Some *status quo* school's characteristics are inputted into the model

to estimate the propensity scores. The inclusion of the baseline characteristics in the probit model is to account for initial conditions, which may influence the ensuing changes in outcome or assignment to the treatment. Hence, ignoring the effects of initial conditions may bias the DID estimates, by wrongly establishing a causal relation between the post adoption of the policy and the outcome variables, when, in fact, the schools could have been enhancing their performance prior to the treatment, and the estimates may be just capturing that trend (Khandker et al., 2009).

To include the initial conditions, I can either lag the variables and insert them differentiated into the probability model - as I did in my main results -, or I can regress the dummy variable for the treatment with the 2007[16] exogenous variables to obtain the propensity scores from the baseline data (see section Sensitivity checks). As such, I choose to follow De Loecker (2007) and lag the variables in respect to the baseline year 2007. This procedure uses the PSM as the estimator to calculate the ATT via non-parametric analysis, whilst, the method of DID-FE uses parametric estimation to estimate the coefficient of treatment. As a result, both results have to be similar to each other, mainly in the signs and directions, as I report in the next sections, which is also a robustness check exercise. After estimating the scores, the next procedure is to check if the PSM is balanced within treated and control groups.

Following Garrido et al. (2014) and Abadie et al. (2004), I divide the sample in n intervals of the same size of the PSM and test for each interval if the score average for the treated corresponds to control. I keep re-dividing the original interval into more blocks and test it again, doing so as many times as necessary to achieve the equivalence between control and treatment groups in all intervals. Eventually, I can test the balance of each confounder, under the null hypothesis that the covariates do not need to be changed, i.e., it is not necessary to re-define the variables included in the probit model, by, for example,

[16]The baseline year for the school entrance age policy is 2007; and for the automatic promotion policy is 2009.

inserting higher order terms or splines to increment the accuracy of the balance. If the null is rejected, then the modification of the covariates is recommended to reach a more accurate balance of scores (Blundell and Costa Dias, 2000).

Using the balanced scores to estimate the probability of starting to adopt the policy, the treated schools are finally matched with their counterparts from the control group. Matching mechanisms abound in the literature, for instance, one-to-one; caliper Cochran and Rubin (1973); nearest neighbour (nn); radius; kernel Heckman et al. (1997, 1998); stratification and Mahalanobis matching Rubin (1980). The latter is useful, in particular, for multiple treatment framework. Caliendo and Kopeinig (2008) explain that any of these matching algorithms may be applied to build the counterfactual. The main difference among them is the way they determine the acceptable neighbourhood - or how the similarity between control and treated is weighted - for the control group to be classified and captured as on common support, thus being considered a match. It follows that each of these approaches present pros and cons that have to be evaluated by the researcher when picking one of them (for more details on the techniques, see (Sianesi, 2001)).

A major issue in non-parametric analysis is the trade-off between bias and variability. For example, in the Kernel matching, each treated school receives a weight and their control counterparts are weighted in conformity with the distance from the treated school, given a certain bandwidth. Hence, the choice of this bandwidth is crucial to the matching procedure, which will be defined by the researcher according to her preference to bias/variance. Thus, on one hand, the larger the chosen range is, the higher the bias will be. On the other hand, a more even density function can be estimated by the kernel approach, representing a more accurate PSM fit, for the reasons given by Garrido et al. (2014) and Caliendo and Kopeinig (2008). In using the Kernel approach, I can mitigate the weight of the contribution to explain the response variable given by the less similar control schools to the treated schools, within a chosen bandwidth of the PSM distribu-

tion. The Kernel estimator smooths the matching curve for the PSM estimation, reducing the median standardised bias between treated and reference groups, and it uses as much information as feasible to do the matching, which make the estimator more precise than the others aforementioned, without losing quality (Sianesi, 2001; Caliendo and Kopeinig, 2008). For this work, as the sample size is large, and as I prefer less bias in my estimates, my bandwidth is 0.01.

The use of PSM-DID is able to mitigate the effect of unobservables that are the same over time. Thus, as long as the researcher controls for school and time fixed effects, she will have to deal just with the confounders that change along the period of the panel, which will be absorbed by the error term. Not being capable of correcting the bias caused by the time-varying unobservables is a limitation of the PSM-DID routine Blundell and Costa Dias (2000), whose PSM-DID estimator is shown below:

$$ATT_{PSM-DID} = \frac{1}{N_T} \sum_{i \in T} [\Delta Y_{i(t+s)} - \sum_{j \in C} w_{ij} \Delta Y_{j(t+s)}] \tag{2.7}$$

where N_T is the number of schools who adopt the new policy in the common support, ΔY_i is the difference between the Mean of the performance of schools before and after adopt the program of school i from the intervention group; ΔY_j denotes the the pre and post difference of school j from the comparison group and w_{ij} is the weight for control school j in the construction of the estimated expected counterfactual result for the treated school i, determined by the score, $\sum_{j \in C} w_{ij} = 1$. (De Loecker, 2007).

The most critical and strong identifying assumption to be met within the DID is the so-called Parallel Trend Assumption (PTA), which needs to hold to guarantee the internal validity of the DID technique. To fulfill this requirement, the difference between the treated and control schools must be constant over time, that is, if by graphic or summary statistics inspection[17], the researcher spots non-parallel trajectories of both control and

[17]Tables A6, A7, A8, A9 show that most of the covariates included in the matching present non-

treatment groups, then it suggests that there are unobservables that vary throughout the panel period that have not been controlled for. That would be the case, for instance, if selection for treatment is influenced by individual-transitory shocks on past outcomes (Ashenfelter's dip[18]) as explained by Abadie (2005).

As the PSM aims solely at removing the bias originated from constant unobservables, as a result, even after the matching procedure, the PTA may be violated as the econometrist cannot control for the former. Therefore, once disregarded the PTA, the estimation of the causality stemming from the treatment effect to the outcome variables will be severely biased (for more details on the PTA, see, e.g., (Angrist and Pischke, 2008; Imbens and Wooldridge, 2009; Card and Krueger, 1993)).

Propensity Score Matching Diagnostic

To assess the rigour of the balancing of the matching applied in this chapter, I follow the tests suggested by the literature on Propensity Score Matching (see e.g. (Austin, 2009; Caliendo and Kopeinig, 2008; Rosenbaum and Rubin, 1985; Blundell and Costa Dias, 2000; Garrido et al., 2014; Sianesi, 2001; Abadie et al., 2004)): i) comparison between the means of the variables inputted in the matching before and after it; ii) observation of the standardized bias for all these variables used to perform the PSM.

The first test consists of comparing the means before and after matching on observables from the treatment and control groups. The expected result for this test is that, after matching, the difference in means between treatment and control groups be statistically equal to zero, i.e, the difference between the two groups should drop substantially after the matching since it is the main purpose of the PSM. The test uses the t-test to

significant difference between treated and control groups after the matching.

[18] "Ashenfelter's dip" is known as the empirical pattern that the Mean earnings of participants in employment and training programs generally decline during the period just before the program (Smith, 1996). "Dip" here Means "trend".

assess the significance of the difference between the means of each variable. It turns out that the customary t-test may lead the researcher to dubious conclusions as Austin (2009) and Garrido et al. (2014) point out. It occurs because the t-test hinges on the sample size, hence, if the number of observations is too high, as it is the case of my sample, the test may indicate non-significance even though the balance of the confounders is not accurate. In this manner, the literature recommends that the second test be used to form the final opinion on the balancing *status* of the covariates.

For the second test, I expect to find the smallest standardised bias suggesting that my matching is balanced, thus, I can use the matched control group to compare with the treated schools. The normalised difference for a continuous variable is:

$$\underbrace{Norm - Diff}_{\text{standardized difference}} = 100 * \frac{\bar{X}_{Tr} - \bar{X}_{Co}}{\sqrt{\frac{S^2_{Tr} + S^2_{Co}}{2}}} \tag{2.8}$$

where $\bar{X}_{Co}$ and $\bar{X}_{Tr}$ are the Mean of each control variable included in the matching for the control and the treated groups; and S^2_{Tr} and S^2_{Co} represent the sample variances for the treated and control schools.

For a categorical variable, the standardised difference is obtained by:

$$\underbrace{Norm - Diff}_{\text{standardized difference}} = 100 * \frac{\hat{J}_{Tr} - \hat{J}_{Co}}{\sqrt{\frac{\hat{J}_{Tr} * (1 - \hat{J}_{Co}) + \hat{J}_{Co} * (1 - \hat{J}_{Co})}{2}}} \tag{2.9}$$

where $\hat{J}_{Tr}$ and $\hat{J}_{Co}$ stand for the Mean of the categorical variable J in treated and control groups.

If the standardised difference between the matched and the unmatched sample is small - under 20% - then the balance achieved is considered good. That being the case, the matched control group can be used to perform the DID technique assuming that group the comparison one to the treated schools. I employ the figure 20% indicated by the seminal work (Rosenbaum and Rubin, 1985), and as the approach used by other researchers

when using PSM, such as Sianesi (2001); Wooldridge (2010); Garrido et al. (2014), among others. The balance tests and the visual balance identification are in Tables A6, A7, A8, A9.

Another step emphasised by Sianesi (2001) when conducting the PSM balancing tests is to check the pseudo-R^2 [19]. I then compare the pseudo-R^2 before and after matching expecting to find a pseudo-R^2 after the PSM smaller than the one found before the procedure. Particularly, a desirable lower pseudo-R^2 will express that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups. Figure 2.1 depicts the observations matched with stars and unmatched with black balls of a PSM execute with the sub-set of municipal schools regarding the treatment automatic promotion. The graph shows that the matched observations are quite close to the vertical line in the zero percent standardised bias, which means that there is very little difference between treatment and control groups after the matching procedure. Furthermore, this plot is representative of a decrease of the pseudo-R^2 from 0.167 before matching to 0.004 after matching, which is the expected result for a good matching.

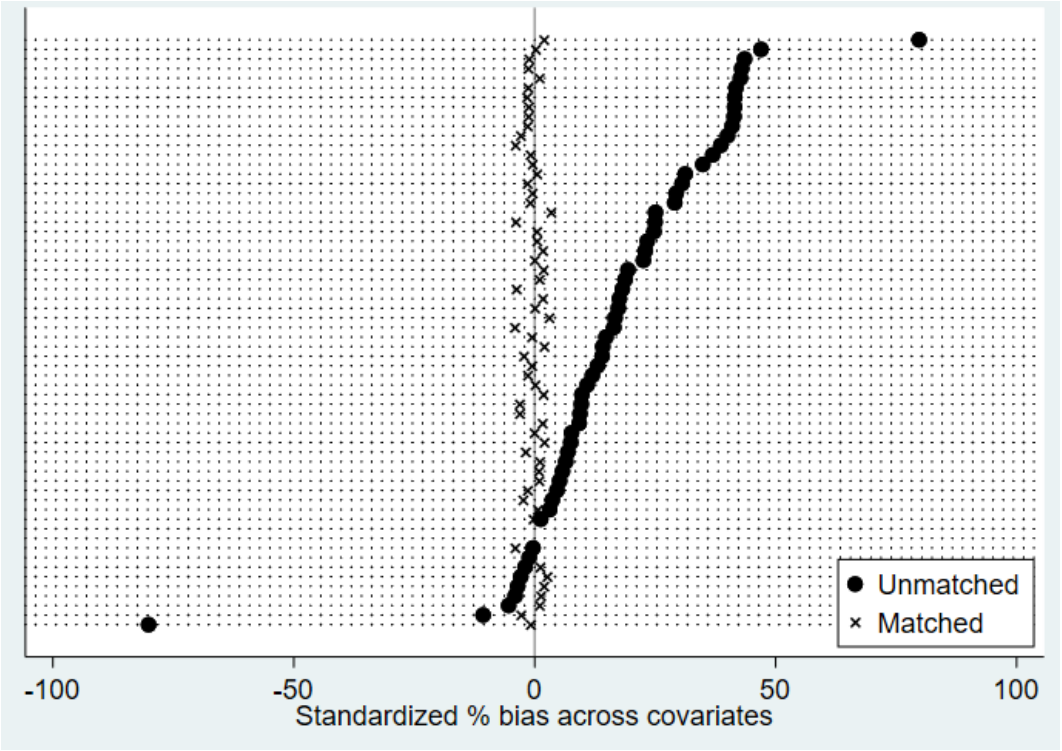
Unobserved Heterogeneity

If the unobserved school heterogeneity persists over time, overlooking it will lead to a serious estimation bias as it would violate the common trend assumption. For instance, a school displayed a particular performance in the *Prova Brasil* examinations in a given year which is due to the influence of a certain unobserved and pre-existent heterogeneity,

[19] The pseudo-R^2 indicates the overall fit of the model. It captures the improvement in the fit of the Maximum likelihood function used for the estimations, when compared to a model that includes just the constant in the regression.

$$Pseudo - R^2 = 1 - \frac{\text{ln likelihood(estimated model)}}{\text{ln likelihood(model with just an intercept)}} \tag{2.10}$$

Figure 2.1: PSM balancing using only municipal schools (primary cycles)



Note: The graph shows that the matched observations are quite close to the vertical line in the zero percent standardised bias, which means that there is very little difference between treatment and control groups after the matching procedure. The pseudo R-square before the matching was 0.167, and after matching, 0.004. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

such as psychological assistance to the families of students who have poor records. Another example related to the case of automatic promotion policy is if schools with high performing pupils and low retention rates adopt non-retention policy, then the response variables and the policy are correlated not only because of the treatment, but also for those other factors, breaching the common trend assumption. If that is the case, the results would be overestimating the actual effects of the treatment on the response variable - Maths and Portuguese scores; pass and dropout rates. In this sense, if I employ standard econometric approaches that do not take into account unobserved characteristics (that do not change over time), I will be neglecting the real effect of the policy (Koppensteiner, 2010).

Therefore, I consider that the error term ν_{it} in Equation 2.11 consists of two compo-

nents, the unobserved school-specific time-invariant effects, k_i, and an exogenous random disturbance, μ_{it}, which gives:

$$\nu_{it} = k_i + \mu_{it} \tag{2.11}$$

where k_i is normally distributed and independent on μ_{it}, and the latter is normally distributed and independent of the independent variables $\forall$ i and t. The unobserved school idiosyncrasies can be controlled either by Fixed Effects (FE) or Random Effects (RE) technique. The former estimates the unobserved characteristics as a parameter, and the latter assumes that it is a random variable. Nonetheless, following Wooldridge (2010) and Vanino (2016), I opted to use the random effects approach when estimating the probability of the schools to adopt the policy since treating the unobserved heterogeneity as a parameter may cause biased estimations of the β_i and of the response variable as a consequence of the incidental parameters problem.

2.5 Empirical Findings

I divide this section into two: the first one presents the results for the compulsory school entrance age policy and the second one covers the findings for the automatic promotion policy.

2.5.1 Compulsory school entrance age

Table 2.7 presents the main estimates of the effect of the adoption of the new compulsory school entrance age on the national examination scores *Prova Brasil*, dropout and pass rates at the school level, for the whole sample. The estimated coefficients refer to the ATT captured by the PSM kernel with bandwidth 0.01. I also conducted the PSM using one to one, nearest neighbour and caliper matching, whose results can be found in the Table A3. The Table presents results for two cohorts. First, cohort 1, in the time $S = 0$ comprises the students who enrolled at the first grade at age 6 in 2007 and took the

exams in 2011 in the fifth grade; cohort 1 in the time $S = 1$ refers roughly[20] to the same students, but who passed the tests in 2015 in the ninth grade. Second, I look at cohort 2, which encompasses the children who started school at the first grade with 6 years old in 2009 and took the fifth grade tests in 2013. For the latter cohort, I cannot examine the second period, $S = 1$ because there is no data available for the exams in 2017, year when they might have taken the exams in the ninth grade. This Table is also divided into set of columns named after letters: letter (a) brings results for Portuguese scores; (b) for Maths; (c) for dropout rates and (d) for pass rates.

From the first set of columns (a), one can see that overall both cohorts increased their Portuguese scores. Cohort 1's raised by 0.3 standard deviation (SD) in $S = 0$ (6.70 marks) and in $S = 1$ (4.79 marks). The second cohort saw an increase of 0.41 SD (10.60 marks). The set of columns (b) shows the scores for Maths. Cohort 1 did not have any statistically significant change in its scores in $S = 0$, but showed an increase of around 0.14 SD (2.41 marks). Cohort 2 shows a rise of 0.29 SD (10.35 marks). Regarding the pass rates, none of them observed significant changes. When it comes to dropout rate, one would expect a decrease in the figures, although what can be noticed is that both cohort 1 in time $S = 1$ and cohort 2 had an increase in the dropout rates: 1.46 pp and 2.61 pp respectively. The small magnitudes of the coefficients are very common to be found in Economics of Education works. For example, Herbst and Strawiński (2016) found a change of 0.11 and 0.04 SD in writing and Maths for the Polish students after the age to enrol at the first grade lowered from 7 to 6. Looking at pupils in Tlaxcal, Mexico, Peña (2017) registered results around 0.3 SD. These works found a decrease in Maths and reading/writing scores when comparing younger children in the same classroom of older ones. My work adds to the literature insofar as it investigates the performance of students in the same classroom with the same age of entrance. It means that I am not comparing the different ages within the

[20] By roughly, I mean the imprecision of the components of the cohort. That is, some students may have failed, changed school and even dropout between 2011 and 2015; hence, it is not necessarily all the same students who took the examinations in 2011, the ones who took in 2015.

Table 2.7: Whole sample - Treatment: compulsory school entrance age - (Kernel)

Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese				(b) Maths			(c) Dropout			(d) Pass		
ATT	6.70***	4.79***	10.60**	3.01	2.41*	10.35***	0.55	1.46*	2.61**	-0.55	1.02	-0.41
	[1.82]	[1.60]	[4.83]	[2.40]	[1.34]	[2.60]	[1.13]	[0.88]	[1.22]	[1.36]	[1.48]	[3.37]
N(control)	306	198	14	306	198	14	308	76	16	308	74	16
N(treat)	16,803	9,978	23,487	16,799	9,978	23,494	22,147	5,922	28,160	22,131	5,910	28,153
N(total)	17,109	10,176	23,501	17,105	10,176	23,508	22,455	5,998	28,176	22,439	5,984	28,169

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

same grade, rather I am looking at similar cohorts (same entrance age) at the school level.

When I split the sample into state and municipal schools (see Tables 2.8 and 2.9), Portuguese scores remain pretty much the same for both types. Whereas, municipal schools are affected by the reform in Maths, pass (a decrease of 4.58 pp) and dropout rates (an increase of 3.96 pp). These results may inform that the change in the compulsory schooling start age to 6 years old made more difficult to students to pass, which can be a good sign that the policy is indeed transforming, even slightly, the quality of education in public schools. Additionally, there is literature on compulsory school start age stressing the lower performance of younger students when compared to the older cohorts due to the lack of maturity and the challenge of starting reading and writing earlier (see the section Literature review for a discussion on this issue). As a result of the increase in the repetition rates, cohort 1 in time $S = 1$ showed a rise in the dropout rate of 3.96 pp. For instance, McEwan and Shapiro (2008) studied a case for Chile and found that a one-year postponement in school admission lowers the probability of failing the first grade by two percentage points; it also increases fourth and eighth grade test scores by more than 0.3 SD, with larger effects for boys[21].

It is worth noting that the pass and dropout rates are collected from the School Census data set, which is a set of information delivered by the schools. On the other hand, the Portuguese and Maths scores are information coming from the results in the *Prova Brasil*, rather than the schools internal examinations. In addition, the signs of Portuguese and Maths scores are positive, and the signs for pass rates are negative, which may appear to be counter-intuitive. It turns out that, the negative sign for pass rates are quite meaningful in accordance with the Brazilian institutions and educational system as the students might be failing the grades because they may be presenting internal tests scores lower than the scores they achieve at the *Prova Brasil*. In other words, the *Prova Brasil* may

[21]For more empirical evidence in support of the magnitude and the signs of the coefficients found in my study, please, see the Literature review section.

Table 2.8: Only municipal schools - Treatment: compulsory school entrance age - (Kernel)

Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese				(b) Maths			(c) Dropout			(d) Pass		
ATT	3.43*	3.60*	11.86**	0.84	1.06	12.78***	1.29	3.96***	1.28	-4.58	1.20	1.1
	[1.96]	[2.03]	[6.09]	[2.21]	[1.62]	[4.14]	[0.94]	[1.68]	[1.16]	[1.37]***	[2.79]	[4.65]
N(control)	155	147	9	155	147	9	157	45	11	157	43	11
N(treat)	26,921	6,449	18,363	26,913	6,449	18,374	32,920	3,463	20,984	32,882	3,434	20,975
N(total)	27,076	6,596	18,372	27,068	6,596	18,383	33,077	3,508	20,995	33,039	3,477	20,986

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate SE by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

be easier than the internal examinations at schools to show to the society that the government is tackling the low quality of education. My sample is from 2007 to 2015, period in which the left-wing party was in the power in Brazil, whose main flag was addressing poverty and education in the country. All in all, the fact that pass rates have decreased throughout the period is another indication that the policy itself is not able to increase the quality of education in Brazil.

Table 2.9: Only state schools - Treatment: compulsory school entrance age - (Kernel)

Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
	(a) Portuguese			(b) Maths			(c) Dropout			(d) Pass		
ATT	5.70***	6.82**	2.26	2.03	2.39	-7.08	-0.29	-0.13	1.99	1.98	2.04	-3.06
	[1.67]	[3.32]	[9.03]	[2.21]	[3.53]	[6.40]	[1.12]	[1.13]	[3.51]	[1.32]	[2.61]	[10.11]
N(control)	275	51	8	275	51	8	277	31	8	277	31	8
N(treat)	6,054	3,497	7,688	6,054	3,497	7,686	8,565	2,238	9,838	8,557	2,255	9,835
N(total)	6,329	3,548	7,696	6,329	3,548	7,694	8,842	2,269	9,846	8,834	2,286	9,843

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

2.5.2 Automatic Promotion

Looking at the second policy analysed in this chapter - automatic promotion - Table 2.10 shows that, for the whole sample, the adoption of automatic promotion increased the Maths scores by 0.19 SD (4.38 marks) for the cohort who started school in 2009 and took the exams in 2013. For the remaining response variables, the reform did not generate any significant effect for those schools that opted for non-retention.

Table 2.10: Whole sample - Treatment: Automatic promotion - (Kernel)

Cohort 1	S=0	S=0	S=0	S=0
Dep. Var.	**(a) Portuguese**	**(b) Maths**	**(c) Dropout**	**(d) Pass**
ATT	1.4	4.38***	-0.1	0.34
	[1.14]	[1.57]	[0.11]	[0.42]
N(control)	4,968	4,968	9,452	9,448
N(treat)	844	844	992	992
N(total)	5,812	5,812	10,444	10,440

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Considering only state schools in Table 2.11, the dropout rate increased by 1.10pp and pass rate by 3.80pp. The latter result is quite intuitive as the policy aim was rising the pass rates to stimulate students not to dropout early from school motivated by failures. It turns out that it seems the policy did not achieve its objective as the dropout rates continued to soar. Moreover, when it comes to only municipal schools sample, Table 2.12 shows that only Maths results had a slight increase for the schools that adopted the non-retention policy (around 0.11 SD).

The literature on automatic promotion is not unanimous. For example, Koppensteiner (2014), investigating only the state of Minas Gerais (one of the wealthiest, located in the Southeast region, which is the better-off in Brazil), found a tiny negative effect of 0.07 SD of the policy on the Maths scores, and non-significant effect on the dropout rates. Concerning the Northeast of Brazil (one of the poorest regions in the country) Gomes-Neto

Table 2.11: Only State schools - Treatment: Automatic promotion - (Kernel)

Cohort 1	S=0	S=0	S=0	S=0
Dep. Var.	**(a) Portuguese**	**(b) Maths**	**(c) Dropout**	**(d) Pass**
ATT	1.34	5.65	1.10***	3.80***
	[9.44]	[16.11]	[0.30]	[1.77]
N(control)	765	765	2,951	2,951
N(treat)	140	140	144	144
N(total)	905	905	3,095	3,095

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

and Hanushek (1994) concluded that Maths scores improved more than Portuguese via grade repetition in that region's public schools. Hence, I can consider that my results are consistent with these empirical works, adding to the literature since I investigate all Brazilian states and municipal schools and found no evidence that passing all students in disregard of whether they are capable to go to the next grade affects the dropout rates and significantly increases literacy and/or the quality of education in Brazil. Other papers in this area show very little or non-significant results regarding the same response variables as mine, indicating that this policy is not the best approach to tackle quality of education and minimise dropout rates either (see for example, (Jacob and Lefgren, 2009; Manacorda, 2012)).

Table 2.12: Only municipal schools - Treatment: Automatic promotion - (Kernel)

Cohort 1	S=0	S=0	S=0	S=0
Dep. Var.	**(a) Portuguese**	**(b) Maths**	**(c) Dropout**	**(d) Pass**
ATT	0.82	2.61*	-0.2488	-0.21
	[0.72]	[1.56]	[0.12]	[0.46]
N(control)	3,931	3,931	6,128	6,124
N(treat)	704	704	848	848
N(total)	4,635	4,635	6,976	6,972

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

2.6 Sensitivity Checks and Limitations

After checking the balance of my PSM, I can merge the "on support" group with the whole sample and run the DID-FE, as detailed in the next sections.

2.6.1 Difference-in-differences with Fixed Effects

The new sample with just on support control and treated schools allows me to proceed to the second part of my methodological strategy, which consists of applying DID-FE to capture the impact of the adoption of the lower school entrance age (or the automatic promotion) on the schools performance in the national examinations *Prova Brasil*, dropout and pass rates. The model of this section uses the DID estimator, and I also control for school fixed effects.

The DID is a quasi-experimental design consisting of, firstly, gauging the difference in the response variables - exams scores, pass and dropout rates - between the period when the schools did not adopt the policy and after they opted by the policy. By comparing the changes in outcomes over time between a population that is enrolled in a program and a population that is not, the DID estimation takes into account the past performance and the idiosyncrasies of the schools in the sample, avoiding potential bias in the estimations stemming from pre-policy characteristics. Secondly, one need to take the difference on the difference calculated for the treated schools minus the control schools. The aftermath is an estimation without the effects of common trends from before the policy adoption.

Following Koppensteiner (2010), under the above identifying assumption, I estimate the effect of the two educational policies discussed in this chapter on the schools' performance in the *Prova Brasil*, dropout and pass rates using the model in Equation 2.12:

$$Y_{st} = \beta_0 + d_s + d_t + \phi d_{st} + \beta_1 X_{st} + \epsilon_{st} \tag{2.12}$$

where Y_{st} is the test score, dropout or pass rate for school s at time t, d_s is a school fixed effects capturing school-specific time invariant effects, d_t is time fixed effects which capture the common time trend of control and treatment groups, d_{st} is the time/treatment interaction term containing information on the treatment status of schools, which varies over time. The coefficient of interest is ϕ, which informs the the average treatment effect of the adoption of the educational reform on the response variables examined here. X_{st} denotes a matrix of exogenous covariates for class and school characteristics - the same used in the PSM model. Therefore, if any difference that lead to the adoption of the policy is included in the school fixed effects, then the parallel trend assumption is not violated.

Bertrand et al. (2004) emphasise the importance of clustering the standard errors at the identification variable level to guarantee that the interaction coefficient is not overestimated. According to the authors, most DID studies rely on years of data and focus on serially correlated outcomes. Nonetheless, most papers disregard the bias in the estimated standard errors that serial correlation introduces. This raises concern since the explanatory variable of interest in DID estimation (e.g. the passage of law) is itself highly serially correlated, producing inflated bias in standard errors. Hence, following the authors, I clustered the DID-FE standard errors per school as Table 2.13 shows.

Table 2.13 depicts roughly the same magnitude and direction of signs as the main results using PSM-DID, corroborating the latter. As already explained in other sections, the DID-FE is a parametric estimation, while the PSM-DID is non-parametric, thus the magnitude difference between the two estimations.

Table 2.13: DID-FE - Treatment: compulsory school entrance age (World Bank Method)

	Whole sample			**State schools**		**Municipal schools**	
	Cohort 1		Cohort 2	Cohort 1		Cohort 1	
Dep.Var./Period	S=0	S=1	S=0	S=0	S=1	S=0	S=1
(a) Portuguese							
Treat(interaction)	3.99***	4.02	25.81	4.23***	9.22*	6.13*	1.88
	[1.14]	[3.45]	[21.13]	[1.28]	[5.41]	[3.82]	[4.71]
N(Obs)	19,266	10,793	16,063	7,525	5,259	11,698	5,348
(b) Maths							
Treat(interaction)	0.84	0.44	19.17	1.41	3.64	0.91	-1.1
	[1.52]	[3.16]	[47.94]	[1.72]	[4.88]	[7.52]	[4.38]
N(Obs)	19,264	10,793	16,067	7,525	5,259	11,696	5,348
(c) Dropout							
Treat(interaction)	0.05	-0.82	6.04**	-0.13	-0.72	1.80***	-0.24
	[0.76]	[1.26]	[2.43]	[0.85]	[1.84]	[0.63]	[1.79]
N(Obs)	25,340	11,366	21,219	10,479	5,493	14,812	5,678
(d) Pass							
Treat(interaction)	1.12	3.45*	-4.14	3.26***	2.14	-3.34	3.65
	[0.92]	[1.82]	[6.57]	[1.00]	[2.71]	[1.69]**	[2.55]
N(Obs)	25,305	11,297	21,217	10,473	5,470	14,804	5,635
School Fixed Effects	yes	yes	yes	yes	yes	yes	yes
Year dummy	yes	yes	yes	yes	yes	yes	yes
Students controls	yes	yes	yes	yes	yes	yes	yes

Notes: Estimations clustered by school, producing robust standard errors. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table 2.14: Treatment: Primary cycles DID-FE (World Bank Method)

	Whole sample	**State schools**	**Municipal schools**
Dependent Variable/Period	S=0	S=0	S=0
(a) Portuguese			
Treat(interaction)	-0.85	-3.04	-0.45
	[0.73]	[3.60]	[0.74]
N(Obs)	8,770	920	7,459
R-squared	0.2110	0.0779	0.2464
(b) Maths			
Treat(interaction)	-2.49	1.64	-2.76
	[1.44]*	[6.20]	[1.52]*
N(Obs)	8,770	921	7,458
R-squared	0.0316	0.0231	0.0479
(c) Dropout			
Treat(interaction)	-2.49	0.31*	-0.18
	[1.44]*	[0.16]	[0.10]*
N(Obs)	8,770	2,752	10,634
R-squared	0.0268	0.0032	0.0355
(d) Pass			
Treat(interaction)	0.73*	-0.72	-1.51
	[0.39]	0.56	[0.26]***
N(Obs)	14,393	2,752	10,634
R-squared	0.0608	0.0050	0.0840
School Fixed Effects	yes	yes	yes
Year dummy	yes	yes	yes
Students controls	yes	yes	yes

Source: Estimations clustered by school, producing robust standard errors. Levels of significance: * 0.10; ** 0.05; *** 0.01.

2.6.2 Other PSM techiques

The main results in this chapter are based on the PSM-DID technique using the Kernel matching with bandwidth 0.01 for the reasons exposed in section 4. In addition to the Kernel type, I also ran Nearest Neighbour, Caliper (0.01) and One to one matching as robustness checks[22]. From the Tables A3, A4, A5, A10, A10 and A12, it is noticeable that the results are consistent with the main Kernel results, across the various specifications.

2.6.3 Placebo Tests

In this section, I perform the PSM-DID estimations based on placebo treatments. From the group of non-treated schools, I draw a random sample of schools - choosing a 10% proportion of the total sample - that I then considered to be the schools that adopted the new school entrance age (or automatic promotion) and implemented the PSM-DID estimator to compare the schools' performance in the *Prova Brasil* exams, the dropout and pass rates to the matched control schools (Huber et al., 2013). I simulated this process 1,000 times and found no significant impacts on the dependent variables (Portuguese and Mathematics scores, dropout and pass rates) caused by the placebo treatment schools as Table 2.15 shows. Therefore, the non-significant results indicate that my main findings displayed in section 5 are not only statistically significant, but also robust.

Concerning the second policy in analysis in this chapter - automatic promotion - I conducted the same placebo treatment test as described above and also found no significant effects of that policy on the outcome variables caused by the placebo treatment (see Table 2.16). Despite the seemingly counter-intuitive significant and negative effect on pass rates, it makes sense as the real treated schools did not present any significant average effect on pass rates due to the policy.

[22]The explanations of each technique can be found in section 4 of this chapter.

Table 2.15: PSM-DID estimations using placebo treatment: compulsory school entrance age

Dep.Var	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1			Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0		S=0	S=1	S=0
	(a) Portuguese			**(b) Maths**			**(c) Dropout**				**(d) Pass**		
ATT	-2.49	-1.55		-3.69	3.9		-0.17	2.34		0	3.65	2.42	
Boostrap S.E.	[3.66]	[5.22]		[4.95]	[4.69]		[2.85]	[2.19]	[1.81]		[2.83]	[3.17]	

Notes: Standard Errors bootstrapped with 1,000 repetitions. I ran 1,000 PSM-DID simulations using placebo treatment. The placebo treatment is not significantly causing the outcome variables from letters a) to d) in the table.

Table 2.16: PSM-DID estimations using placebo treatment: automatic promotion

Cohort 1	S=0	S=0	S=0	S=0
	(a) Portuguese	**(b) Maths**	**(c) Dropout**	**(d) Pass**
ATT	0.77	1.21	0.08	-0.7***
Boostrap S.E.	[0.86]	[1.45]	[0.08]	[0.25]

Notes: Standard Errors bootstrapped with 1,000 repetitions. I ran 1,000 PSM-DID simulations using placebo treatment. The placebo treatment is not significantly causing the outcome variables from letters a) to d) in the table.

2.6.4 Limitations

This work does not follow an experimental research design as I analyse one policy which is totally optional (the automatic promotion); and although the second one (the compulsory school entrance age) is compulsory and reaches all the public and private schools, I look at the optional phase because there is not enough time lapsed since the beginning of the compulsory phase (in 2010), results and data set to analyse it. In spite of the fact that randomisation cannot be guaranteed, there are similarities between control and treated groups because I just kept public schools in my sample and controlled for a plethora of schools, students and staff features. Moreover, the remaining differences between the two groups were overcome by the PSM method, persisting just generally minimal and relatively stable over time differences. Hence, I can sustain a plausible causal inference between the treatments and the outcomes of my research (Murnane and Willett, 2010).

Even though the data set on education in Brazil is resourceful and detailed, there are some observed and unobserved variables and some observations missing from the different data sets that I merged to build the panel data for this chapter. That may influence the final estimations of student dropout and pass rates, as well as the scores. Thus, the estimated effect of the policies may justifiably attract some critical reviews.

2.7 Conclusion

In this chapter, I analyse the effect of two independent educational policies on the schools performance at the national examination *Prova Brasil*, dropout and pass rates. The first policy made it optional for schools to adopt the automatic promotion from the first to the ninth grades. In 2000, a massive number of schools started to introduce the non-retention policy. I found no evidence of the impact of this policy on the Portuguese and Maths' scores at the *Prova Brasil*. In addition, I found significant and positive coefficients for pass rates, which is the aim of the practice of non-retention. However, the results indicate also a positive relationship between non-repetition in the primary school and dropout rates, which may be signalling that the policy itself does not suffice to change the alarming dropout scenario and to keep students away from the streets.

The second policy increased the compulsory years of schooling from 8 to 9 years, by anticipating the school starting age from 7 to 6 years old. The law had an optional phase, from 2006-2010, offering the schools the opportunity to choose whether they would accept younger children in the first grade. This non-mandatory phase was due to the necessity of important changes in the infra-structure, staff training and didactic resources. From 2010 onwards, all the schools were obliged to welcome children at the age 6 in the first grade. My results indicate that this policy raised the Portuguese and Maths scores in the national examinations, but also increased the dropout rates and lowered the pass rates. The latter is likely to happen to do the fact that the schools internal examinations are harder them the *Prova Brasil* due to political issues discussed in the previous section.

All in all, these results suggest that it is not only by issuing new educational rules without attempting to the states, municipalities, schools and children idiosyncrasies that the country will enhance the quality of education. The stakeholders particularities of a vast Brazil must be taken into consideration; interdisciplinary and contextualised methodologies must ensure meaningful learning for students. Therefore, new profiles of staff, training

them, monitoring not only the students, but also the schools, and the professionals are of the utmost importance to improve the quality of education in Brazil. These remarks can be applied to other developing countries as well, with governments ensuring adequate availability of complementary tools to influence the accomplishment of the policies objectives, the efficiency and quality of primary education.

Future research can cast light on the compulsory SSA policy phase, when there is data set available to assess the causality of the reform on the response variables analysed in this paper. Furthermore, other causal relations can be considered; for example, crime in Brazil or in one specific state in Brazil and the SSA, via the incapacitating effect stemming from more years of schooling.

Appendix

Table A1: Marks scale for Portuguese exam in *Prova Brasil*

	Portuguese		
Level	5th grade	9th grade	high school
Up to level 1	0 - 149 marks	———	———
Level 1	———	200 - 224 marks	225 - 249 marks
Level 2	150 - 174 marks	225 - 249 marks	250 - 274 marks
Level 3	175 - 199 marks	250 - 274 marks	275 - 299 marks
Level 4	200 - 224 marks	275 - 299 marks	300 - 324 marks
Level 5	225 - 249 marks	300 - 324 marks	325 - 349 marks
Level 6	250 - 274 marks	325 - 349 marks	350 - 374 marks
Level 7	275 - 299 marks	350 - 374 marks	375 - 399 marks
Level 8	300 - 324 marks	375 - 400 marks	400 - 425 marks
Level 9	325 - 350 marks	———	———

Source: Brazilian Ministry of Education.SAEB (2006)

Table A2: Marks scale for Mathematics exam in *Prova Brasil*

Mathematics			
Level	5th grade	9th grade	high school
Level 1	125 - 149 marks	200 - 224 marks	225 - 249 marks
Level 2	150 - 174 marks	225 – 249 marks	250 - 274 marks
Level 3	175 - 199 marks	250 - 274 marks	275 - 299 marks
Level 4	200 - 224 marks	275 - 299 marks	300 - 324 marks
Level 5	225 - 249 marks	300 - 324 marks	325 - 349 marks
Level 6	250 - 274 marks	325 - 349 marks	350 - 374 marks
Level 7	275 - 299 marks	350 - 374 marks	375 - 399 marks
Level 8	300 - 324 marks	375 - 399 marks	400 - 424 marks
Level 9	325 - 350 marks	400 - 425 marks	425 - 449 marks
Level 10	————	————	450 - 475 marks

Source: Brazilian Ministry of Education. SAEB (2006)

Table A6: Matching diagnostic - treatment: compulsory school entrance age (Kernel method); cohort 1; s=0; only STATE SCHOOLS)

	Unmatched	Mean			Mean %reduct	t-test	$p > \|t\|$
Variable	Matched	Treated	Control	%bias	bias		
dpropfemale	U	0.92831	1.9712	-13.9		-2.15	0.032
	M	0.92831	1.6466	-9.6	31.1	-5.44	0
dpropwhite	U	-0.17382	-1.4408	14.2		2.18	0.029
	M	-0.17382	-1.3655	13.3	5.9	7.26	0
dpropbrown	U	-0.72852	-0.1894	-4.8		-0.75	0.453
	M	-0.72852	0.10378	-7.4	-54.4	-3.55	0
dpropblack	U	-0.77739	-1.6363	13.4		2.34	0.019
	M	-0.77739	-1.76	15.4	-14.4	7.62	0
dpropyellow	U	-0.98631	-0.37012	-21.4		-3.31	0.001
	M	-0.98631	-0.96679	-0.7	96.8	-0.36	0.72
dpropindigenous	U	-1.4394	-1.4328	-0.2		-0.03	0.974
	M	-1.4394	-1.1963	-7.2	-3593.9	-4.07	0

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dpropeightyrs_less	U	-0.76388	-1.3848	21.7		3.18	0.002
	M	-0.76388	-0.66527	-3.4	84.1	-2.12	0.034
dpropnineyrs	U	-11.859	-19.827	54.9		9.31	0
	M	-11.859	-13.932	14.3	74	8.06	0
dproptenyrs	U	5.5919	11.312	-43.7		-7.49	0
	M	5.5919	5.8478	-2	95.5	-1.25	0.21
dpropelevenyrs	U	7.8412	8.7959	-8.7		-1.17	0.242
	M	7.8412	9.3635	-13.8	-59.5	-7.5	0
dproptwelveyrs	U	1.2465	2.7547	-27.7		-4.49	0
	M	1.2465	1.0235	4.1	85.2	2.05	0.041
dpropthirteenyrs	U	0.20856	0.6095	-10.9		-1.77	0.076
	M	0.20856	0.36111	-4.1	62	-2.26	0.024
dpropfourteenyrs	U	-0.08314	0.14031	-9.2		-1.56	0.118
	M	-0.08314	-0.09567	0.5	94.4	0.3	0.767
dpropfifteenyrs_more	U	-0.3381	-0.08443	-11.3		-1.78	0.076
	M	-0.3381	-0.26856	-3.1	72.6	-1.77	0.076
dpropworkyes	U	5.9833	16.081	-91.8		-15.42	0
	M	5.9833	7.3125	-12.1	86.8	-6.63	0
dproppresch	U	15.61	32.77	-92.8		-14.26	0
	M	15.61	22.937	-39.6	57.3	-21.96	0
dpropfailed	U	-11.133	-13.376	19.9		3.01	0.003
	M	-11.133	-13.275	19	4.5	10.79	0
dproplivemother	U	3.7864	3.7395	0.8		0.12	0.907
	M	3.7864	4.1609	-6.2	-698.2	-3.27	0.001
dproplivefather	U	3.7138	3.9838	-3.1		-0.5	0.619
	M	3.7138	4.363	-7.5	-140.4	-3.95	0

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dpropnevermothersch	U	-7.4029	-6.5145	-15		-2.43	0.015
	M	-7.4029	-9.7375	39.4	-162.8	18.08	0
dpropnoneight	U	-13.119	-9.863	-44		-6.77	0
	M	-13.119	-14.408	17.4	60.4	9.16	0
dpropnonhighschool	U	1.8287	-0.67503	29.3		4.7	0
	M	1.8287	3.7125	-22	24.8	-12.36	0
dpropparents_tell_study	U	7.5907	8.8665	-14.5		-2.1	0.036
	M	7.5907	6.8245	8.7	39.9	5.07	0
dpropparents_homework_incentive	U	3.724	2.9012	14.1		2.03	0.043
	M	3.724	3.6667	1	93	0.52	0.601
dpropparents_read_incentive	U	4.8825	3.1708	27		3.99	0
	M	4.8825	4.0264	13.5	50	8.03	0
dpropattendclass_incentive	U	5.4107	4.9936	6.9		1.03	0.305
	M	5.4107	5.5581	-2.5	64.7	-1.37	0.172
sch_building	U	1.0499	1.0655	-5		-0.83	0.407
	M	1.0499	1.0614	-3.7	26.4	-2.19	0.028
filtered_water	U	1.4685	1.3745	16.5		2.83	0.005
	M	1.4685	1.4678	0.1	99.3	0.07	0.945
water_provision	U	0.96829	0.93091	17.1		3.38	0.001
	M	0.96829	0.89554	33.3	-94.6	16.06	0
sewer_public	U	0.80624	0.67273	30.7		5.43	0
	M	0.80624	0.64436	37.3	-21.2	20.29	0
sewer_fossa	U	0.22762	0.36727	-30.9		-5.36	0
	M	0.22762	0.41338	-41.1	-33	-22.35	0
garbage_collection	U	0.98563	0.98545	0.1		0.02	0.981
	M	0.98563	0.98765	-1.7	-1053.4	-0.97	0.334

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
principal_room	U	0.95821	0.95636	0.9		0.15	0.881
	M	0.95821	0.97618	-8.9	-873.8	-5.56	0
teachers_room	U	0.93921	0.94182	-1.1		-0.18	0.86
	M	0.93921	0.95487	-6.6	-501.2	-3.85	0
computer_lab	U	0.89924	0.94909	-18.9		-2.71	0.007
	M	0.89924	0.91566	-6.2	67.1	-3.12	0.002
science_lab	U	0.16881	0.36364	-45.1		-8.32	0
	M	0.16881	0.11341	12.8	71.6	8.78	0
special_care_room	U	0.17906	0.25818	-19.2		-3.32	0.001
	M	0.17906	0.14673	7.8	59.2	4.82	0
sports_pitch	U	0.75752	0.74909	2		0.32	0.75
	M	0.75752	0.6697	20.4	-942.3	10.74	0
kitchen	U	0.97704	0.99636	-16.9		-2.13	0.033
	M	0.97704	0.97332	3.3	80.7	1.32	0.188
library	U	0.67228	0.75636	-18.7		-2.92	0.004
	M	0.67228	0.64411	6.3	66.5	3.27	0.001
park	U	0.055	0.02545	15.1		2.13	0.034
	M	0.055	0.0374	9	40.4	4.62	0
dbathroom_outdoor	U	0.01536	-0.00727	7.6		1.56	0.118
	M	0.01536	0.05141	-12.1	-59.3	-7.12	0
bathroom_specialneeds	U	0.26924	0.32364	-11.9		-1.98	0.047
	M	0.26924	0.25683	2.7	77.2	1.55	0.121
special_needs_room	U	0.19739	0.26909	-17		-2.91	0.004
	M	0.19739	0.20124	-0.9	94.6	-0.53	0.596
dn_rooms	U	0.01652	-0.52364	16.5		2.36	0.018
	M	0.01652	-0.11733	4.1	75.2	2.28	0.023

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
tv	U	0.99686	0.98909	9.3		2.14	0.032
	M	0.99686	0.9868	12	-29.4	6.16	0
videocassete	U	0.8373	0.82545	3.2		0.52	0.603
	M	0.8373	0.80972	7.4	-132.9	3.98	0
dvd	U	0.97225	0.97818	-3.8		-0.59	0.556
	M	0.97225	0.98189	-6.2	-62.5	-3.54	0
parabolica	U	0.69343	0.67636	3.7		0.6	0.549
	M	0.69343	0.62402	14.9	-306.8	8.07	0
xerox	U	0.5261	0.48	9.2		1.5	0.134
	M	0.5261	0.48974	7.3	21.1	4	0
retroproj	U	0.8411	0.89818	-17		-2.55	0.011
	M	0.8411	0.90199	-18.1	-6.7	-10.05	0
printer	U	0.9848	0.97455	7.3		1.34	0.18
	M	0.9848	0.98905	-3	58.6	-2.06	0.04
dn_pcs	U	7.4836	4.9091	13.4		1.65	0.098
	M	7.4836	6.7855	3.6	72.9	2.04	0.041
dadm_pcs	U	1.6929	1.1273	17.3		2.64	0.008
	M	1.6929	1.8646	-5.3	69.7	-3.12	0.002
dstud_pcs	U	5.1027	3.9491	6.3		0.77	0.439
	M	5.1027	4.0782	5.6	11.2	3.15	0.002
internet	U	0.92666	0.87273	18		3.31	0.001
	M	0.92666	0.92329	1.1	93.8	0.7	0.482
dn_staff	U	1.3168	0.05818	7.7		1.22	0.223
	M	1.3168	1.521	-1.2	83.8	-0.64	0.519
sch_meal	U	0.98959	0.98182	6.5		1.22	0.221

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	0.98959	0.99032	-0.6	90.6	-0.4	0.687

Notes: Reported are the means of variables for treated and control schools for unmatched and matched sample together with the bias and standardized differences (% bias reduction) and t-tests in matched sample compared to those in unmatched sample. Pseudo r2 before matching was 0.256; after matching, 0.172. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A3: All the PSM techniques used

	One to one			**Nearest neighbour**			**Caliper**			**Kernel**		
Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese												
ATT	7.12***	4.46**	0.23	8.56***	6.09***	11.02*	9.96***	4.36	0.53	6.70***	4.79***	10.60**
	[2.46]	[2.15]	[6.15]	[3.23]	[2.53]	[5.90]	[2.91]	[2.95]	[9.45]	[1.82]	[1.60]	[4.83]
N(control on support)	306	198	14	306	198	14	306	198	14	306	198	14
N(treat on support)	13,009	8,498	1,971	16,803	9,978	23,487	14,822	9,962	23,467	16,803	9,978	23,487
N(total)	17,109	10,176	23,501	17,109	10,176	23,501	17,109	10,160	23,501	17,109	10,176	23,501
(b) Maths												
ATT	4.83*	2.86	5.79*	8.57***	2.63	12.82***	9.25***	2.53	10.24**	3.01	2.41*	10.35***
	[2.90]	[1.81]	[3.29]	[4.26]	[2.12]	[3.17]	[3.92]	[2.48]	[4.90]	[2.40]	[1.34]	[2.60]
N(control on support)	306	198	14	306	198	14	306	198	14	306	198	14
N(treat on support)	13,006	8,498	1,982	16,799	9,978	23,494	14,816	9,962	23,474	16,799	9,978	23,494
N(total)	17,105	10,176	23,494	17,105	10,176	23,508	17,105	10,160	23,508	17,105	10,176	23,508
(c) Dropout												
ATT	1.64	0.76	-0.51	1.34	0.03	0.5	2.14	1.45	-0.42	0.55	1.46*	2.61**
	[1.88]	[1.31]	[2.86]	[2.48]	[1.50]	[2.26]	[2.32]	[1.44]	[2.42]	[1.13]	[0.88]	[1.22]
N(control on support)	308	76	16	308	76	16	308	76	16	308	76	16
N(treat on support)	15,660	4,587	8,733	22,147	5,922	28,160	18,933	5,906	28,127	22,147	5,922	28,160
N(total)	22,455	5,922	28,176	22,455	5,998	28,176	22,455	5,998	28,160	22,455	5,998	28,176
(d) Pass												
ATT	-5.73	-0.05	1.21	-2.11	2.43	0.15	-3.04	2.8	2.66	-0.55	1.02	-0.41
	[2.59]**	[2.84]	[8.12]	[3.00]	[2.62]	[6.23]	[2.75]	[2.82]	[6.65]	[1.36]	[1.48]	[3.37]
N(control on support)	308	76	16	308	74	16	308	74	16	308	74	16
N(treat on support)	15,692	4,587	8,760	22,131	5,910	28,153	18,922	5,883	28,120	22,131	5,910	28,153
N(total)	22,415	5,998	28,119	22,439	5,984	28,169	22,439	5,984	28,136	22,439	5,984	28,169

Notes: Estimations using bootstrap with 500 repetitions. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A4: Propensity Score Matching - Difference in Differences (only states schools)

	One to one			**Nearest Neighbour**			**Caliper**			**Kernel**		
Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese												
ATT	5.2***	6.36**	7.88	4.66**	7.77	12.61*	4.59***	3.16	-4.99	5.70***	6.82**	2.26
	[1.97]	[3.16]	[6.51]	[2.40]	[4.76]	[7.36]	[2.05]	[5.19]	[14.11]	[1.67]	[3.32]	[9.03]
N(control on support)	275	51	8	275	51	8	275	51	8	275	51	8
N(treat on support)	5,105	1,468	190	6,054	3,497	7,688	5,255	3,479	7,464	6,054	3,497	7,688
N(total)	6,329	3,548	7,696	6,329	3,548	7,696	6,329	3,548	7,696	6,329	3,548	7,696
(b) Maths												
ATT	1.32	4.84	-2.13	1.25	7.39	-2.15	1.10	3.65	-10.61	2.03	2.39	-7.08
	[2.52]	[3.42]	[5.33]	[3.19]	[5.07]	[5.21]	[2.69]	[5.81]	[10.35]	[2.21]	[3.53]	[6.40]
N(control on support)	275	51	8	275	51	8	275	51	8	275	51	8
N(treat on support)	5,107	1,468	190	6,054	3,497	7,686	5,257	3,497	7,644	6,054	3,497	7,686
N(total)	6,054	3,548	7,686	6,329	3,548	7,694	6,329	3,548	7,694	6,329	3,548	7,694
(c) Dropout												
ATT	-0.11	-0.79	3.66	-0.57	0.03	5.77**	-0.32	0.66	-0.14	-0.29	-0.13	1.99
	[1.49]	[0.86]	[2.37]	[1.90]	[1.50]	[2.55]	[1.38]	[1.57]	[4.28]	[1.12]	[1.13]	[3.51]
N(control on support)	277	31	8	277	76	8	277	31	8	277	31	8
N(treat on support)	7,100	621	206	8,565	5,922	9,838	7,136	2,216	9,764	8,565	2,238	9,838
N(total)	8,842	2,269	9,846	8,842	5,998	9,846	8,842	2,269	9,846	8,842	2,269	9,846
(d) Pass												
ATT	0.69	-0.66	-1.16	0.56	1.46	-5.64	1.90	2.46	-1.74	1.98	2.04	-3.06
	[1.99]	[2.59]	[7.39]	[2.23]	[3.05]	[8.07]	[1.51]	[4.06]	[13.24]	[1.32]	[2.61]	[10.11]
N(control on support)	277	31	8	277	31	8	277	31	8	277	31	8
N(treat on support)	7,069	621	206	8,557	2,255	9,835	7,133	2,218	9,791	8,557	2,255	9,835
N(total)	8,808	2,269	9,843	8,834	2,286	9,843	8,834	2,286	9,843	8,834	2,286	9,843

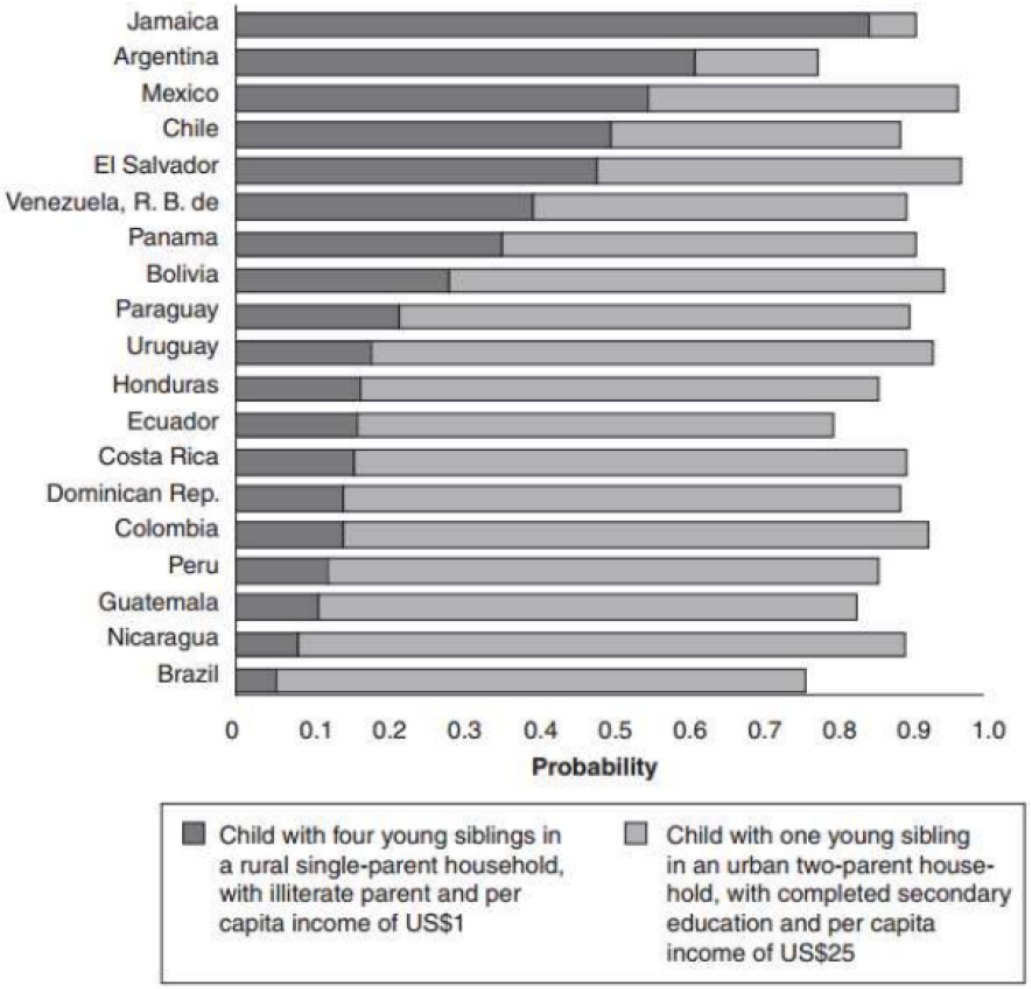
Notes: Estimations using bootstrap with 500 repetitions. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A5: Propensity Score Matching - Difference in Differences (only municipal schools)

	One to one			**Nearest Neighbour**			**Caliper**			**Kernel**		
Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese												
ATT	4.64	2.83	5.33	2.63	5.46*	10.01	3.8	8.09***	-0.29	3.43*	3.60*	11.86**
	[4.24]	[2.54]	[4	[4.20]	[3.05]	[7.57]	[4.27]	[3.40]	[13.05]	[1.96]	[2.03]	[6.09]
N(control on support)	155	147	14,780	155	147	9	155	147	9	155	147	9
N(treat on support)	24,167	5,363	7	26,921	6,449	18,363	26,896	6,419	18,315	26,921	6,449	18,363
N(total)	27,076	6,596	155	27,076	6,596	18,372	27,076	6,596	18,372	27,076	6,596	18,372
(b) Maths												
ATT	5.32	1.47	13.37***	2.53	1.92	13.05***	3.82	2.45	10	0.84	1.06	12.78***
	[4.75]	[2.02]	[4.18]	[4.70]	[2.43]	[5.13]	[4.79]	[2.72]	[8.90]	[2.21]	[1.62]	[4.14]
N(control on support)	155	147	7	155	147	9	155	147	9	155	147	9
N(treat on support)	24,158	5,363	155	26,913	6,449	18,374	26,889	6,419	18,326	26,913	6,449	18,374
N(total)	27,068	6,449	14,951	27,068	6,596	18,383	27,068	6,596	18,383	27,068	6,596	18,383
(c) Dropout												
ATT	-0.58	0.31	2.64*	-0.65	0.03	3.17*	-0.50	9.98***	3.48	1.29	3.96***	1.28
	[2.06]	[1.39]	[1.56]	[2.21]	[1.50]	[1.83]	[2.11]	[3.98]	[3.10]	[0.94]	[1.68]	[1.16]
N(control on support)	157	45	9	157	76	11	157	45	11	157	45	11
N(treat on support)	28,328	1,112	557	32,920	5,998	20,984	32,892	3,446	20,964	32,920	3,463	20,984
N(total)	33,077	3,508	17,364	33,077	5,998	20,995	33,077	3,508	20,995	33,077	3,508	20,995
(d) Pass												
ATT	1.23	2.46	-8.29	-4.02	1.68	-2.26	-3.97	5.39	-8.02	-4.58	1.20	1.1
	[3.80]	[2.05]	[3.41]***	[3.17]	[4.31]	[7.36]	[3.13]	[5.92]	[13.08]	[1.37]***	[2.79]	[4.65]
N(control on support)	157	43	9	157	43	11	157	43	11	157	43	11
N(treat on support)	28,303	1,097	557	32,882	3,434	20,975	32,840	3,406	20,995	32,882	3,434	20,975
N(total)	33,049	3,477	17,348	33,039	3,477	20,986	33,039	3,477	20,986	33,039	3,477	20,986

Notes: Estimations using bootstrapping to calculate standard errors (SE) with 500 repetitions. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Figure A1: Simulated probability of completion sixth grade on time, *circa* 2005



Source: (Barros, 2015). Note: Brazil legs behind most of its Latin American neighbours when it comes to schooling completion and dropping out probabilities.

Table A7: Matching diagnostic - treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; whole sample)

	Unmatched	Mean			Mean	%reduct	t-test	$p > \|t\|$
Variable	Matched	Treated	Control	%bias	bias			
white	U	-6.0955	-3.3689	-23.70			-2.99	0.003
	M	-6.0955	-4.88	-10.50		55.4	-7.17	0
brown	U	0.50467	-2.5569	20.50			2.75	0.006

Continued on next page

Table A7 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	0.50467	-1.0806	10.60	48.2	7.76	0
black	U	-1.2365	-1.2567	0.30		0.04	0.967
	M	-1.2365	-0.7404	-7.60	-2358	-5.27	0
asian	U	-1.0448	-1.3074	7.9	7.90	1.06	0.29
	M	-1.0448	-0.92288	-3.7	53.6	-2.64	0.008
indigenous	U	-0.92646	-1.2471	9.2		1.22	0.223
	M	-0.92646	-1.1159	5.4	40.9	4.09	0
8 or less years old	U	-1.9752	-2.0168	1.3		0.17	0.865
	M	-1.9752	-2.0692	3	-125.7	2.13	0.033
9 years old	U	-20.924	-17.48	-32.8		-4.46	0
	M	-20.924	-18.674	-21.4	34.7	-14.68	0
10 years old	U	-3.4681	-1.758	-11.4		-1.61	0.107
	M	-3.4681	-3.1116	-2.4	79.1	-1.64	0.101
11 years old	U	26.315	23.715	17.8		2.31	0.021
	M	26.315	25.728	4	77.4	2.75	0.006
12 years old	U	2.2103	1.9681	3.4		0.48	0.632
	M	2.2103	1.1062	15.6	-356	10.95	0
13 years old	U	0.52037	0.07584	8.4		1.31	0.191
	M	0.52037	0.3824	2.6	69.2	1.89	0.058
14 years old	U	0.20957	-0.7809	31.1		4.28	0
	M	0.20957	0.00775	6.3	79.6	4.64	0
15 or more years old	U	-0.60228	-1.0389	11.3		1.66	0.096
	M	-0.60228	-1.1512	14.3	-25.7	10.3	0
worker	U	-3.8841	-3.0412	-8.5		-1.23	0.01
	M	-3.8841	-4.3495	4.7	44.8	3.56	0
attended preschool	U	-2.7417	0.61704	-19.1		-2.58	0.01

Continued on next page

Table A7 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	-2.7417	-0.4522	-13	31.8	-9.11	0
ever failed	U	-7.0169	-6.4208	-4.6		-0.64	0.524
	M	-7.0169	-7.8083	6.2	-32.8	4.45	0
live with mother	U	5.7703	3.9142	16		1.9	0.057
	M	5.7703	4.5399	10.6	33.7	7.45	0
live with father	U	0.50045	-0.49298	8.4		1.06	0.289
	M	0.50045	-0.76266	10.6	-27.1	7.33	0
mother never school	U	-9.001	-10.695	23.8		3.53	0
	M	-9.001	-10.086	15.2	36	11.24	0
mother less 8th grade	U	-13.381	-11.308	-23.3		-3.14	0.002
	M	-13.381	-12.974	-4.6	80.4	-3.3	0.001
mother less high school	U	-1.4628	-0.80645	-7.4		-1.03	0.304
	M	-1.4628	-2.0476	6.6	10.9	4.64	0
parents tell to study	U	6.8404	7.5854	-8.6		-1.16	0.245
	M	6.8404	6.7677	0.8	90.2	0.62	0.537
parents homework incentive	U	5.3361	3.4125	25.2		3.2	0.001
	M	5.3361	4.5839	9.8	60.9	6.94	0
parents read incentive	U	7.4657	5.7551	19		2.3	0.021
	M	7.4657	6.2937	13	31.5	9.24	0
paernts class incentive	U	5.9425	3.0418	32.3		4.01	0
	M	5.9425	4.8555	12.1	62.5	9.34	0
school building	U	0.97284	0.95455	9.8		1.56	0.119
	M	0.97284	0.97499	-1.1	88.2	-0.95	0.341
filtered water	U	0.75476	0.90909	-42.1		-5.02	0
	M	0.75476	0.84373	-24.3	42.4	-15.78	0
kitchen	U	0.98637	0.98485	1.3		0.18	0.855

Continued on next page

Table A7 – continued from previous page

Variable	Unmatched Matched	Mean Treated	Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	0.98637	0.98941	-2.6	-100	-1.96	0.049
library	U	0.8181	0.73737	19.5		2.91	0.004
	M	0.8181	0.78944	6.9	64.5	5.1	0
park	U	0.3168	0.10101	55		6.5	0
	M	0.3168	0.21148	26.8	51.2	16.99	0
bathroom outdoor	U	0.01824	-0.0202	12.2		1.71	0.088
	M	0.01824	0.03197	-4.3	64.3	-2.81	0.005
bathroom special needs	U	0.47334	0.44444	5.8		0.81	0.42
	M	0.47334	0.47028	0.6	89.4	0.43	0.665
special needs room	U	0.4082	0.36869	8.1		1.12	0.263
	M	0.4082	0.43715	-5.9	26.7	-4.14	0
number of rooms	U	0.48908	0.26263	4.7		0.78	0.436
	M	0.48908	0.39727	1.9	59.4	1.03	0.304
tv	U	0.98878	0.98485	3.4		0.52	0.605
	M	0.98878	0.98582	2.6	24.8	1.86	0.063
videocassete	U	0.6367	0.55556	16.6		2.35	0.019
	M	0.6367	0.61186	5.1	69.4	3.62	0
dvd	U	0.93325	0.89394	14		2.18	0.029
	M	0.93325	0.92841	1.7	87.7	1.35	0.177
parabolic	U	0.39707	0.44444	-9.6		-1.35	0.178
	M	0.39707	0.38931	1.6	83.6	1.12	0.261
xerox	U	0.67769	0.51515	33.5		4.84	0
	M	0.67769	0.65556	4.6	86.4	3.32	0.001
overhead projector	U	0.77931	0.70707	16.6		2.42	0.015
	M	0.77931	0.74457	8	51.9	5.77	0
printer	U	0.94488	0.93434	4.4		0.64	0.521

Continued on next page

Table A7 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	0.94488	0.93986	2.1	52.3	1.52	0.128
number of pcs	U	13.775	12.485	4.5		0.48	0.634
	M	13.775	13.9	-0.4	90.3	-0.31	0.759
staff members	U	4.2504	5.9697	-7.2		-1.12	0.263
	M	4.2504	3.6294	2.6	63.9	1.82	0.069
school meal	U	0.97555	0.95455	11.4		1.88	0.06
	M	0.97555	0.95825	9.4	17.6	6.84	0.000

Notes: Reported are the means of variables for treated and control schools for unmatched and matched sample together with the bias and standardized differences (% bias reduction) and t-tests in matched sample compared to those in unmatched sample. Pseudo R-squared unmatched: 0.083; after matching: 0.060. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A8: Matching diagnostic - treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; only state schools)

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dpropwhite	U	-9.8676	-4.4294	-44.5		-2.95	0.003
	M	-9.8676	-15.88	49.2	-10.6	18.37	0
dpropbrown	U	0.33063	0.29438	0.3		0.02	0.986
	M	0.33063	4.9814	-35.2	12728.5	-13.22	0
dpropblack	U	-1.5229	-1.5119	-0.2		-0.01	0.99
	M	-1.5229	-0.22016	-22.6	11761.8	-8.62	0

Continued on next page

Table A8 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dpropyellow	U	-1.1765	-1.8419	19.8		1.41	0.158
	M	-1.1765	-2.4886	39.1	-97.2	18.93	0
dpropindigenous	U	-0.80761	-1.5868	25.2		1.64	0.101
	M	-0.80761	0.11438	-29.8	-18.3	-13.36	0
dpropeightyrs_less	U	-2.3408	-1.7515	-21		-1.48	0.14
	M	-2.3408	-3.1575	29.1	-38.6	8.81	0
dpropnineyrs	U	-26.301	-23.545	-27.5		-1.84	0.066
	M	-26.301	-19.818	-64.7	-135.3	-28.76	0
dproptenyrs	U	-4.312	-5.0022	4.1		0.32	0.747
	M	-4.312	-5.9572	9.7	-138.4	4.32	0
dpropelevenyrs	U	30.131	29.004	7.3		0.5	0.615
	M	30.131	30.873	-4.8	34.1	-2.33	0.02
dproptwelveyrs	U	2.6934	3.0587	-5.6		-0.38	0.702
	M	2.6934	1.5072	18.3	-224.7	7.86	0
dpropthirteenyrs	U	0.87114	1.152	-6.1		-0.47	0.641
	M	0.87114	-0.06298	20.4	-232.5	10.91	0
dpropfourteenyrs	U	0.44007	-0.5832	38.8		2.59	0.01
	M	0.44007	-1.9526	90.7	-133.8	34.07	0
dpropfifteenyrs_more	U	-0.03611	0.35363	-19.2		-1.13	0.259
	M	-0.03611	-0.1098	3.6	81.1	1.6	0.109
dpropworkyes	U	-4.5475	-1.1249	-36.4		-2.66	0.008
	M	-4.5475	-5.2647	7.6	79	3.35	0.001
dproppresch	U	-6.053	-3.0195	-18		-1.22	0.223
	M	-6.053	1.8636	-47	-161	-21.2	0
dpropfailed	U	-9.0934	-5.1382	-35		-2.33	0.02
	M	-9.0934	-11.75	23.5	32.8	9.49	0

Continued on next page

Table A8 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dproplivemother	U	1.2183	1.7773	-7		-0.58	0.561
	M	1.2183	-0.19432	17.6	-152.7	9.18	0
dproplivefather	U	-2.0537	-2.4981	4.2		0.29	0.769
	M	-2.0537	-8.118	57.7	-1264.7	22.62	0
dpropnevermothersch	U	-7.0631	-6.3135	-14		-0.98	0.329
	M	-7.0631	-10.38	62.1	-342.5	24.53	0
dpropnoneight	U	-13.933	-12.855	-11.3		-0.82	0.411
	M	-13.933	-18.911	52.3	-361.9	20.72	0
dpropnonhighschool	U	-3.4496	-4.8231	17.5		1.15	0.252
	M	-3.4496	-0.32133	-39.9	-127.8	-14.23	0
dpropparents_tell_study	U	3.8795	8.9173	-61.4		-5.67	0
	M	3.8795	3.3114	6.9	88.7	3.71	0
dpropparents_homework_incentive	U	3.6223	5.0511	-25.8		-1.84	0.066
	M	3.6223	4.7604	-20.5	20.3	-10.05	0
dpropparents_read_incentive	U	4.9377	7.0108	-28.7		-2.28	0.023
	M	4.9377	4.3967	7.5	73.9	3.22	0.001
dpropattendclass_incentive	U	3.184	4.5575	-24.4		-1.74	0.081
	M	3.184	3.665	-8.5	65	-4.41	0
sch_building	U	0.97998	0.92157	27		2.9	0.004
	M	0.97998	0.98952	-4.4	83.7	-3.25	0.001
filtered_water	U	0.60967	0.86275	-59.7		-3.69	0
	M	0.60967	0.69057	-19.1	68	-7.12	0
library	U	0.89105	0.84314	14.1		1.09	0.277
	M	0.89105	0.88104	2.9	79.1	1.32	0.188
park	U	0.39291	0.09804	72.7		4.3	0
	M	0.39291	0.28161	27.4	62.3	9.91	0

Continued on next page

Table A8 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dbathroom_outdoor	U	0.00515	-0.03922	17.1		1.02	0.307
	M	0.00515	-0.0024	2.9	83	1.43	0.154
bathroom_specialneeds	U	0.47984	0.39216	17.7		1.24	0.213
	M	0.47984	0.45057	5.9	66.6	2.45	0.014
special_needs_room	U	0.41607	0.39216	4.9		0.34	0.731
	M	0.41607	0.4114	0.9	80.5	0.4	0.692
dn_rooms	U	0.10237	-0.88235	31.3		2.25	0.024
	M	0.10237	0.26719	-5.2	83.3	-2.64	0.008
tv	U	0.99457	0.96078	22.8		3.13	0.002
	M	0.99457	0.99914	-3.1	86.5	-3.42	0.001
videocassete	U	0.70403	0.62745	16.2		1.19	0.235
	M	0.70403	0.67367	6.4	60.3	2.74	0.006
dvd	U	0.94309	0.96078	-8.2		-0.54	0.588
	M	0.94309	0.9949	-24.1	-192.9	-12.64	0
parabolica	U	0.43752	0.5098	-14.4		-1.03	0.302
	M	0.43752	0.32249	23	-59.1	9.98	0
xerox	U	0.78381	0.56863	47		3.69	0
	M	0.78381	0.91246	-28.1	40.2	-15.23	0
retroproj	U	0.87046	0.88235	-3.6		-0.25	0.802
	M	0.87046	0.9735	-31.2	-766.4	-16.37	0
printer	U	0.97312	0.98039	-4.8		-0.32	0.75
	M	0.97312	0.99644	-15.4	-220.7	-8	0
dn_pcs	U	15.959	14.078	13.8		1.02	0.309
	M	15.959	11.079	35.8	-159.5	17.15	0
dn_staff	U	2.0309	12.549	-36		-3.88	0
	M	2.0309	5.03	-10.3	71.5	-6.19	0

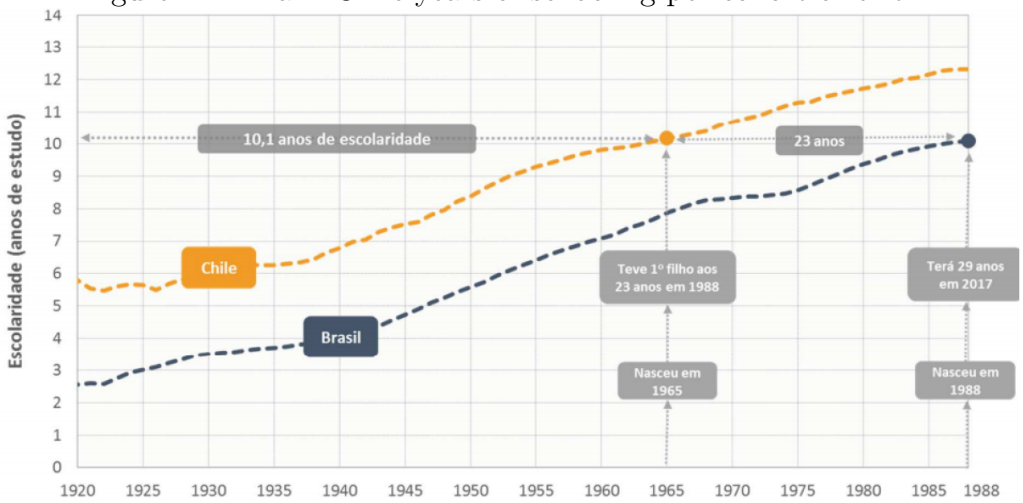
Continued on next page

Table A8 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
sch_meal	U	0.97226	0.96078	6.3		0.49	0.621
	M	0.97226	0.99233	-11.1	-74.9	-6.38	0

Notes: Reported are the means of variables for treated and control schools for unmatched and matched sample together with the bias and standardized differences (% bias reduction) and t-tests in matched sample compared to those in unmatched sample. Pseudo r2 from 0.237 to 0.485 after matching. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Figure A2: Brazil-Chile years of schooling per cohort of birth



Source: (Barros, 2015). Note: The level of schooling (ordinate in the graph) of Brazilian young adults (29 years old) is the same as the Chilean youngsters' parents when they were in the same age (29 years old) 23 years ago. That is, Brazil is lagging one generation behind.

Table A9: Matching diagnostic - Treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; only municipal schools)

	Unmatched	Mean			%reduct		
Variable	Matched	Treated	Control	%bias	bias	t-test	$p > \|t\|$
dpropwhite	U	-4.0583	-3.001	-9.6		-1.05	0.293
	M	-4.0583	-4.2476	1.7	82.1	0.94	0.349
dpropbrown	U	0.62304	-3.5462	26.9		3.11	0.002
	M	0.62304	-1.0864	11	59	6.26	0
dpropblack	U	-1.0789	-1.1681	1.3		0.15	0.881
	M	-1.0789	-0.59497	-7	-442.8	-3.96	0
dpropyellow	U	-0.97639	-1.1219	4.4		0.5	0.62
	M	-0.97639	-0.83878	-4.1	5.5	-2.31	0.021
dpropindigenous	U	-0.99325	-1.1293	3.7		0.43	0.669
	M	-0.99325	-1.1935	5.5	-47.2	3.29	0.001
dpropeightyrs_less	U	-1.7712	-2.1089	10.2		1.11	0.269

Continued on next page

Table A9 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	-1.7712	-1.9745	6.2	39.8	3.69	0
dpropnineyrs	U	-17.967	-15.376	-26.9		-3.23	0.001
	M	-17.967	-16.099	-19.4	27.9	-10.5	0
dproptenyrs	U	-3.0201	-0.63327	-16.9		-1.97	0.049
	M	-3.0201	-3.257	1.7	90.1	0.93	0.352
dpropelevenyrs	U	24.223	21.88	16.9		1.85	0.065
	M	24.223	24.041	1.3	92.2	0.7	0.484
dproptwelveyrs	U	1.9451	1.5898	4.9		0.59	0.553
	M	1.9451	1.0032	12.9	-165.1	7.26	0
dpropthirteenyrs	U	0.33005	-0.29754	11.3		1.51	0.13
	M	0.33005	0.42591	-1.7	84.7	-0.97	0.33
dpropfourteenyrs	U	0.08568	-0.8495	27.6		3.27	0.001
	M	0.08568	0.1537	-2	92.7	-1.2	0.232
dpropfifteenyrs_more	U	-0.91598	-1.522	14		1.76	0.079
	M	-0.91598	-1.7105	18.4	-31.1	11.07	0
dpropworkyes	U	-3.5257	-3.7061	1.8		0.22	0.824
	M	-3.5257	-4.5028	9.6	-441.6	5.89	0
dproppresch	U	-0.96542	1.8787	-16.1		-1.89	0.059
	M	-0.96542	1.7517	-15.4	4.5	-8.43	0
dpropfailed	U	-5.8834	-6.8658	7.4		0.88	0.38
	M	-5.8834	-7.4106	11.5	-55.5	6.65	0
dproplivemother	U	8.258	4.6556	28.1		2.77	0.006
	M	8.258	5.5965	20.7	26.1	11.84	0
dproplivefather	U	1.8912	0.20266	13.6		1.45	0.147
	M	1.8912	0.82235	8.6	36.7	4.91	0
dpropnevermothersch	U	-10.069	-12.216	29.2		3.67	0

Continued on next page

Table A9 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	-10.069	-11.328	17.1	41.3	10.18	0
dpropnoneight	U	-13.074	-10.772	-26.6		-3.03	0.002
	M	-13.074	-11.791	-14.8	44.3	-8.73	0
dpropnonhighschool	U	-0.36581	0.58706	-10.6		-1.28	0.199
	M	-0.36581	-1.42	11.8	-10.6	6.89	0
dpropparents_tell_study	U	8.4474	7.1232	15		1.64	0.102
	M	8.4474	7.5317	10.4	30.8	6.08	0
dpropparents_homework_incentive	U	6.2709	2.844	40.9		4.36	0
	M	6.2709	4.7088	18.6	54.4	10.48	0
dpropparents_read_incentive	U	8.8472	5.3194	36.2		3.61	0
	M	8.8472	6.4413	24.7	31.8	14.08	0
dpropattendclass_incentive	U	7.4451	2.5159	49.1		5.13	0
	M	7.4451	5.2303	22.1	55.1	13.71	0
sch_building	U	0.96899	0.96599	1.7		0.21	0.836
	M	0.96899	0.97408	-2.9	-69.8	-1.74	0.082
filtered_water	U	0.83362	0.92517	-28.4		-2.96	0.003
	M	0.83362	0.87765	-13.6	51.9	-7.13	0
kitchen	U	0.98387	0.97959	3.2		0.41	0.685
	M	0.98387	0.98902	-3.8	-20.2	-2.53	0.011
library	U	0.77857	0.70068	17.8		2.24	0.025
	M	0.77857	0.76401	3.3	81.3	1.97	0.049
park	U	0.27508	0.10204	45.3		4.67	0
	M	0.27508	0.21012	17	62.5	8.63	0
dbathroom_outdoor	U	0.02528	-0.01361	11.6		1.47	0.142
	M	0.02528	0.06499	-11.9	-2.1	-5.86	0
bathroom_specialneeds	U	0.47124	0.46259	1.7		0.21	0.835

Continued on next page

Table A9 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	0.47124	0.47387	-0.5	69.6	-0.3	0.765
special_needs_room	U	0.40487	0.36054	9.1		1.08	0.279
	M	0.40487	0.43732	-6.7	26.8	-3.73	0
dn_rooms	U	0.6984	0.65986	0.7		0.1	0.918
	M	0.6984	1.0151	-6	-721.7	-2.67	0.008
tv	U	0.98573	0.9932	-7.3		-0.76	0.448
	M	0.98573	0.98509	0.6	91.3	0.31	0.759
videocassete	U	0.59978	0.53061	14		1.69	0.091
	M	0.59978	0.57845	4.3	69.2	2.46	0.014
dvd	U	0.92821	0.87075	19.2		2.65	0.008
	M	0.92821	0.92028	2.6	86.2	1.7	0.089
parabolica	U	0.37556	0.42177	-9.4		-1.14	0.253
	M	0.37556	0.36079	3	68	1.74	0.082
xerox	U	0.62025	0.4966	25.1		3.05	0.002
	M	0.62025	0.63153	-2.3	90.9	-1.32	0.186
retroproj	U	0.72911	0.64626	17.9		2.23	0.026
	M	0.72911	0.71935	2.1	88.2	1.24	0.215
printer	U	0.92929	0.91837	4.1		0.51	0.61
	M	0.92929	0.91396	5.8	-40.3	3.24	0.001
dn_pcs	U	12.579	11.932	1.9		0.17	0.866
	M	12.579	13.367	-2.3	-21.7	-1.31	0.192
dn_staff	U	5.3901	3.6871	7.9		0.95	0.342
	M	5.3901	4.6528	3.4	56.7	1.92	0.055
sch_meal	U	0.97767	0.95238	13.8		2.03	0.043
	M	0.97767	0.96158	8.8	36.4	5.33	0

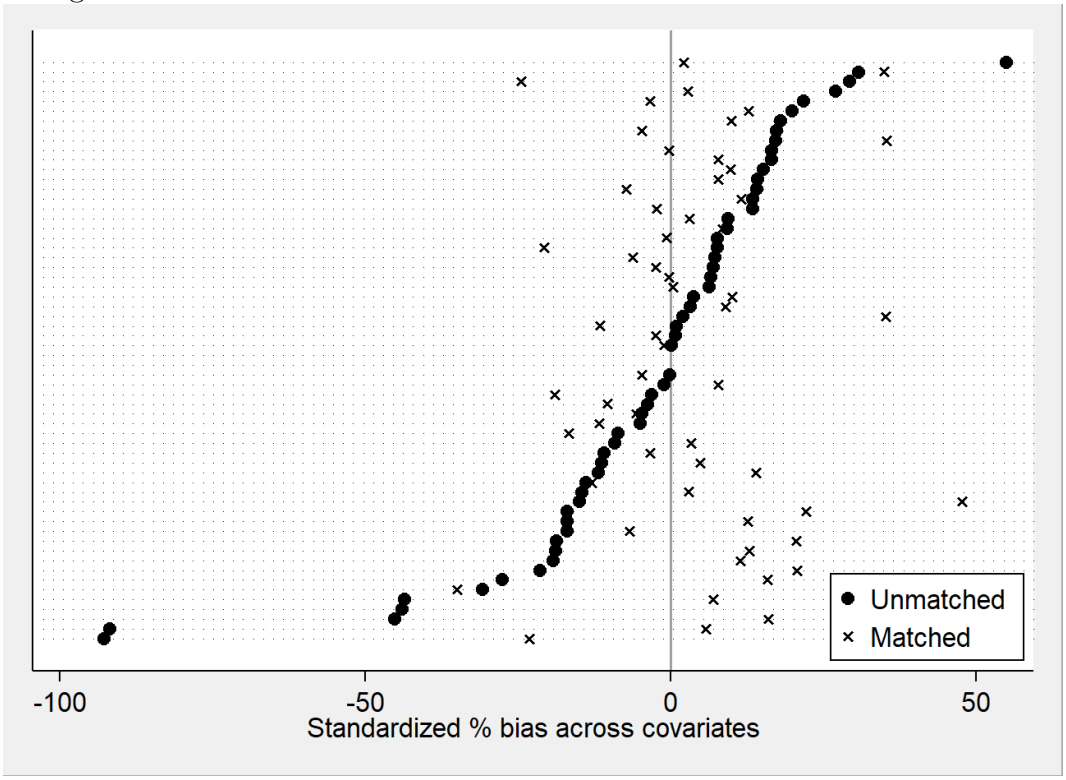
Continued on next page

Table A9 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test $p > \|t\|$

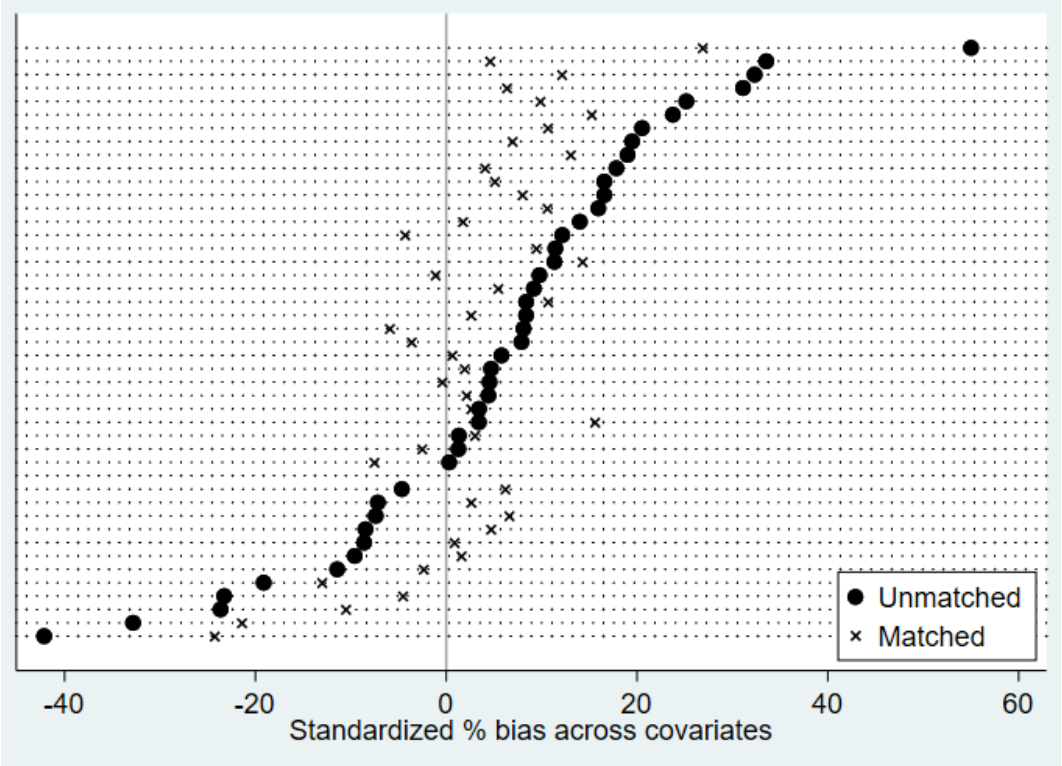
Notes: Reported are the means of variables for treated and control schools for unmatched and matched sample together with the bias and standardized differences (% bias reduction) and t-tests in matched sample compared to those in unmatched sample. Pseudo r2 from 0.088 to 0.080 after matching. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

Figure A3: PSM balancing - state schools only - Method: Kernel - compulsory school entrance age



Notes: The pseudo R-square before the matching was 0.256, and after matching, 0.004. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

Figure A4: PSM balancing - whole sample s=1 - Kernel - compulsory school entrance age



The pseudo R-square before the matching was 0.256, and after matching, 0.004. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

The political-legal education system reforms timeline SEB/MEC (2006):

- Law no. 4,024, of December 20, 1996 - Established 4 years of Elementary School.
- Punta del Este and Santiago Agreement - Commitment to establish six years for Teaching Fundamental until 1970.
- Law no. 5.692, of August 11, 1971 - Obligation of Elementary School of eight years.
- Law 9,394, of December 20, 1996 - admits the enrolment in Primary Education of nine years, the six years of age.
- Law no. 10. 172, of January 9, 2001.
  - Approved the National Education Plan / PNE.
  - A nine-year primary has become the progressive goal of national education.
- Law no. 11. 114, of May 16, 2005 - makes it compulsory for the children of six years of age to age in Elementary School.

Table A10: Whole Sample - Kernel - automatic promotion

	One to one	**NN**	**Caliper**	**Kernel**
Dep. Var.	Cohort 1	Cohort 1	Cohort 1	Cohort 1
	S=0	S=0	S=0	S=0
(a) Portuguese				
ATT	1.4	1.31	0.58	1.29*
	[1.14]	[0.89]	[1.18]	[0.76]
N(control on support)	4,968	4,968	4,968	4,968
N(treat on support)	844	844	836	844
N(total)	5,812	5,812	5,812	5,812
(b) Maths				
ATT	1.54	3.65**	3.73*	4.38***
	[2.37]	[1.92]	[2.28]	[1.57]
N(control on support)	4,968	4,968	4,968	4,968
N(treat on support)	844	844	836	844
N(total)	5,812	5,812	5,804	5,812
(c) Dropout				
ATT	-0.09	-0.14	-0.03	-0.1
	[0.15]	[0.12]	[0.15]	[0.11]
N(control on support)	9,452	9,452	9,452	9,452
N(treat on support)	992	992	986	992
N(total)	10,444	10,444	10,444	10,444
(d) Pass				
ATT	-0.004	0.51	0.19	0.34
	[0.57]	[0.47]	[0.54]	[0.42]
N(control on support)	9,448	9,448	9,448	9,448
N(treat on support)	992	992	985	992
N(total)	10,440	10,440	10,440	10,440

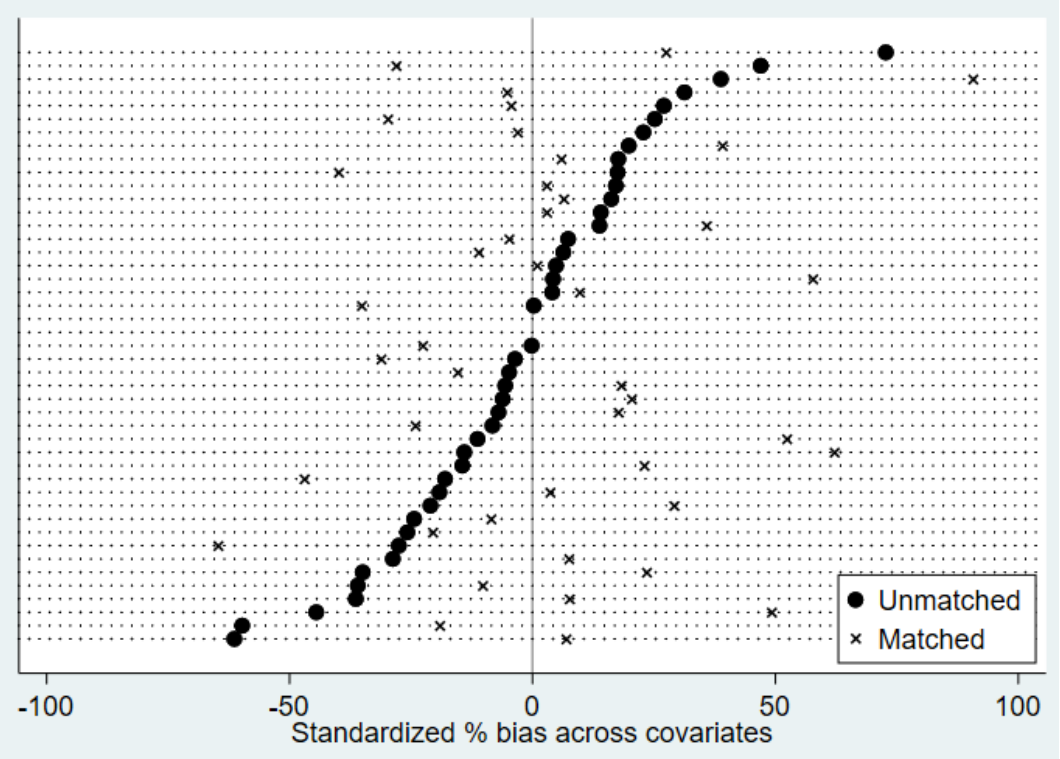
Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works. I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM (I used 500 repetitions). Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A11: State schools - Kernel - automatic promotion

	One to one	**NN**	**Caliper**	**Kernel**
Dep. Var.	Cohort 1	Cohort 1	Cohort 1	Cohort 1
	S=0	S=0	S=0	S=0
(a) Portuguese				
ATT	3.91	4.22	3.03	1.34
	[5.77]	[10.09]	[7.35]	[9.44]
N(control on support)	765	765	765	765
N(treat on support)	94	140	71	140
N(total)	905	905	905	905
(b) Maths				
ATT	7.92	15.97	-1.72	5.65
	[11.10]	[18.10]	[9.73]	[16.11]
N(control on support)	765	765	765	765
N(treat on support)	94	140	71	140
N(total)	905	905	905	905
(c) Dropout				
ATT	0.75	1.43***	1.18***	1.10***
	[0.48]	[0.51]	[0.48]	[0.30]
N(control on support)	2,951	2,951	2,951	2,951
N(treat on support)	136	144	96	144
N(total)	3,095	3,095	3,095	3,095
(d) Pass				
ATT	3.48***	2.55	2.72	3.80***
	[1.63]	[1.61]	[1.69]	[1.77]
N(control on support)	2,951	2,951	2,951	2,951
N(treat on support)	136	144	96	144
N(total)	3,095	3095	3,095	3,095

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works. I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM (I used 500 repetitions). Levels of significance: * 0.10; ** 0.05; *** 0.01.

Figure A5: PSM balancing -state schools s=1 - Kernel - compulsory school entrance age



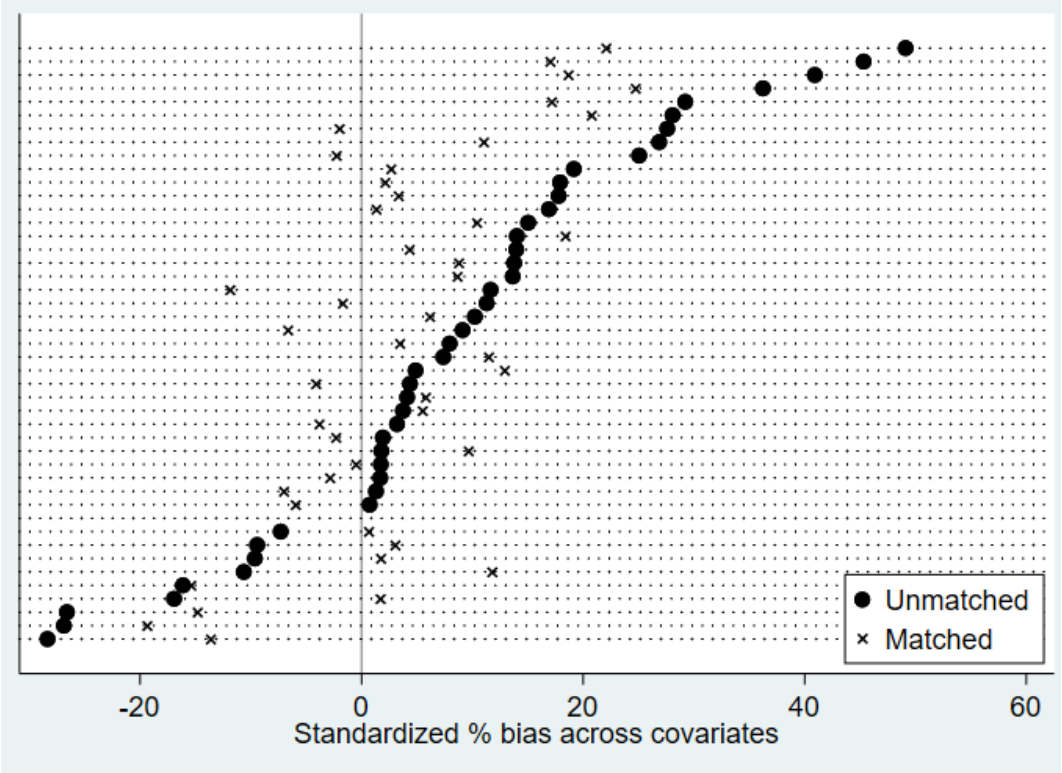
Note: The pseudo R-square before the matching was 0.580, and after matching, 0.223. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

- Law no. 11.274, of February 6, 2006 - extends the Elementary School for a nine-year term, with the enrolment of six-year-old children and establishes a systems by 2010.

Variables for matching and control:

1. **Students characteristics per school**:
    - gender: percentage of students per gender;
    - age: proportion of average age per school;
    - pre-school: percentage of students who attended pre-school;
    - race: proportion of students per race;
    - work: percentage of students who work;
    - percentage of students who ever failed a grade;
    - percentage of students who live with mother/father;

Figure A6: PSM balancing - municipal s=1 - Kernel - compulsory school entrance age



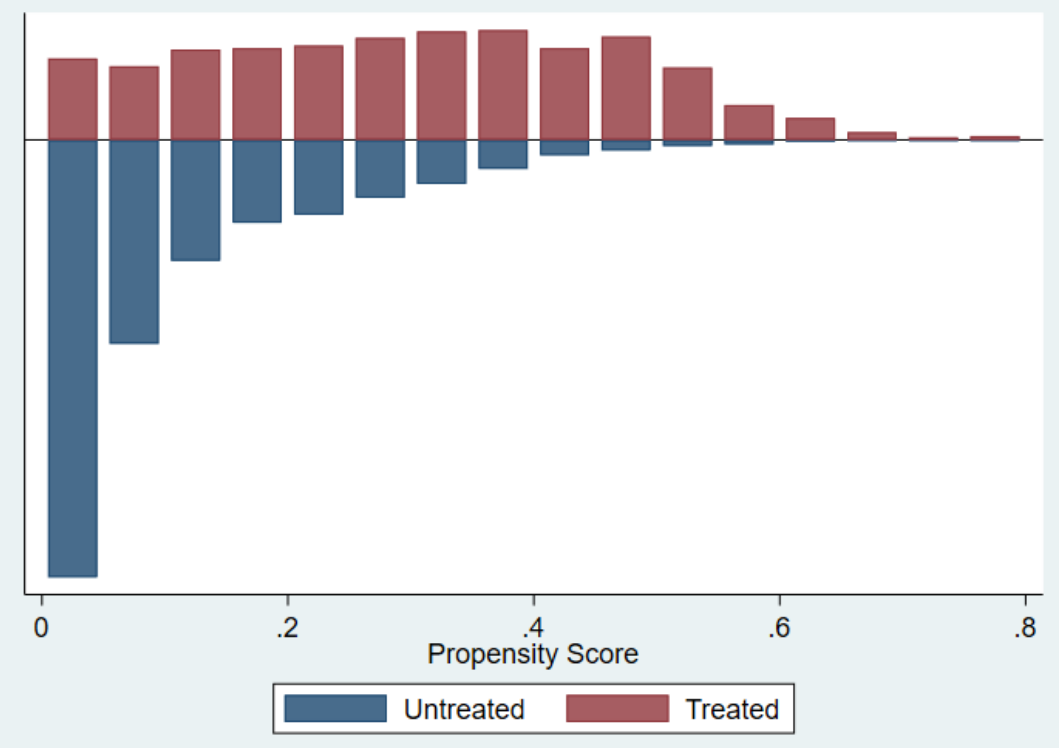
The pseudo R-square before the matching was 0.467, and after matching, 0.214. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

- mother schooling
- parents' interest in children education:
  - parents' incentive to attend the classes;
  - parents' incentive to do the homework;
  - parents' incentive to read;
  - parents' incentive to study.

2. **Schools characteristics**:

- infrastructure:
  - numbers of computers per student;
  - internet access;
  - science laboratory;
  - computer laboratory;
  - teachers room

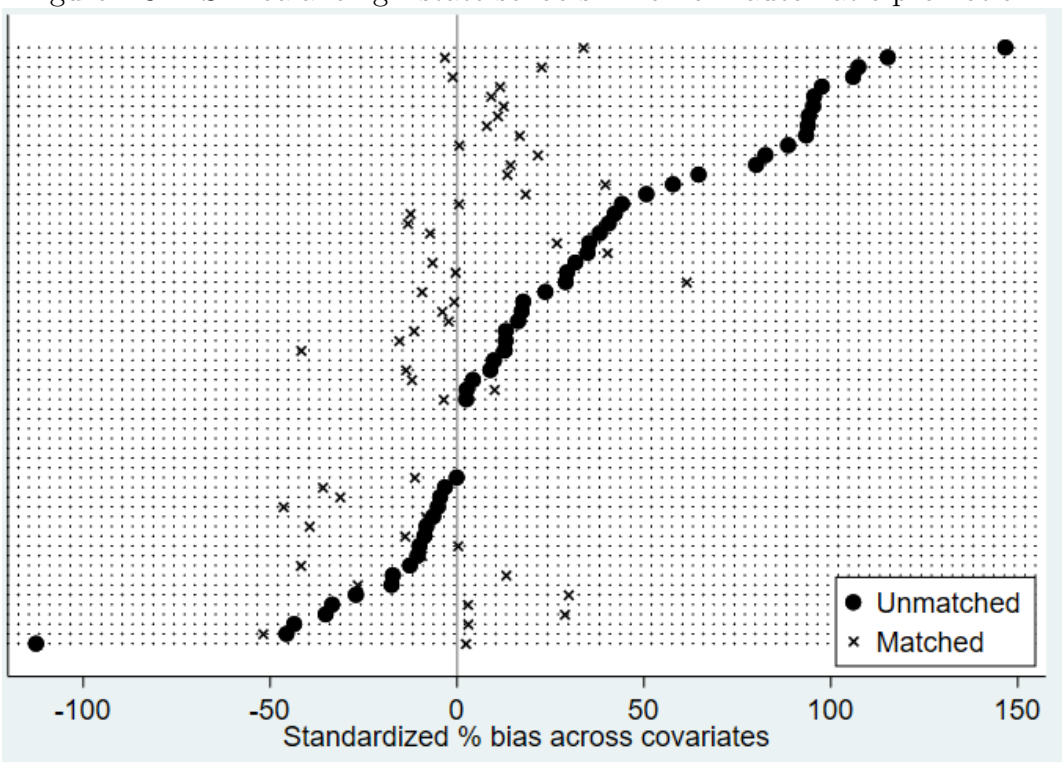
Figure A7: PSM balancing - whole sample - Kernel - automatic promotion)



Notes: The pseudo R-square before the matching was 0.204, and after matching, 0.008. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

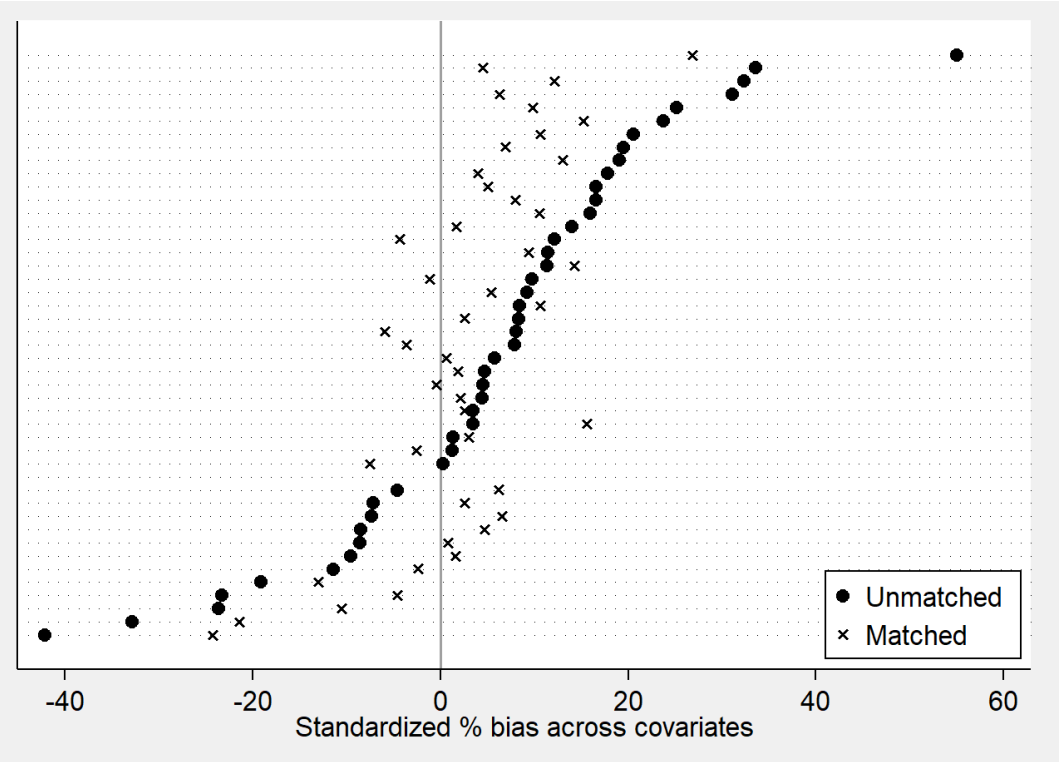
- quality of education:
  - percentage of teachers with university diploma

Figure A8: PSM balancing - state schools - Kernel - automatic promotion



Notes: The pseudo R-square before the matching was 0.629, and after matching, 0.371. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

Figure A9: PSM diagnostic balancing - whole sample s= 1 - Kernel - compulsory school entrance age



The pseudo R-square before the matching was 0.730, and after matching, 0.391. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

Table A12: Municipal schools - Kernel - automatic promotion

	One to one	**NN**	**Caliper**	**Kernel**
Dep. Var.	Cohort 1	Cohort 1	Cohort 1	Cohort 1
	S=0	S=0	S=0	S=0
(a) Portuguese				
ATT	0.92	0.67	1.33	0.82
	[0.96]	[0.82]	[1.10]	[0.72]
N(control on support)	3,931	3,931	3,931	3,931
N(treat on support)	702	704	701	704
N(total)	4,635	4,635	4,635	4,635
(b) Maths				
ATT	1.3	1.85	2.78	2.61*
	[2.24]	[1.84]	[2.26]	[1.56]
N(control on support)	3,931	3,931	3,931	3,931
N(treat on support)	702	704	701	704
N(total)	4,635	4,635	4,635	4,635
(c) Dropout				
ATT	-0.2	-0.11	-0.12	-0.24
	[0.16]	[0.14]	[0.17]	[0.12]**
N(control on support)	6,128	6,128	6,128	6,128
N(treat on support)	847	848	839	848
N(total)	6,975	6,976	6,976	6,976
(d) Pass				
ATT	-0.2	-0.4	-0.39	-0.21
	[0.62]	[0.51]	[0.59]	[0.46]
N(control on support)	6,124	6,124	6,124	6,124
N(treat on support)	847	848	839	848
N(total)	6,972	6,972	6,972	6,972

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works. I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM (I used 500 repetitions). Levels of significance: * 0.10; ** 0.05; *** 0.01.

Chapter 3

The effects of air pollution on students' cognitive performance at Brazilian university entrance tests: Evidence from changes in Wind Direction

3.1 Introduction

The World Health Organisation (WHO) estimates that globally as many as 4.2 million premature deaths each year are linked to exposure to ambient air pollution.[1] The WHO also points out that the health impacts associated with pollution exposure are not equitably distributed, with the poorest individuals in the poorest countries being disproportionately affected. The manner in which many common local air pollutants adversely affect health, particularly via the respiratory and cardiovascular systems, is now well known (Dockery et al., 1993; Hansen and Selte, 2000; Weinhold, 2008; Currie et al., 2009;

[1] http://www9.who.int/airpollution/ambient/health-impacts/en/

Ling and van Eeden, 2009; Currie et al., 2009; Neidell, 2009; Developers, 2016). However, a more recent literature has examined the possibility that exposure to air pollution may also affect cognitive performance and, in turn, potentially human capital formation, productivity, and ultimately economic growth (Zweig et al., 2009; Currie et al., 2014; Ebenstein et al., 2016). If true, this raises the prospect of a 'cycle of disadvantage' in which exposure to pollution constrains the development prospects of those who need it most.

The literature examining the impact of pollution exposure on human capital formation has found compelling evidence that air pollution, through its impact on health, increases school absenteeism (Currie et al., 2009; Ransom and Pope, 2013; Chen et al., 2018; Liu and Salvo, 2018), negatively affects school pupils' scores (Almond and Edlund, 2009; Zweig et al., 2009; Nilsson, 2009; Reyes, 2011; Sanders, 2012; Bharadwaj et al., 2017), and decreases low and high-skilled workers' productivity (Chang et al., 2016; Heyes et al., 2016; Archsmith et al., 2018; Chang et al., 2019; Kahn and Li, 2019). Other studies cast light on the accumulated impact of air pollution within a period of time on pupils' scores at school (Zweig et al., 2009). Overall, these studies look at the long-run impact of fetal (Almond and Edlund, 2009; Sanders, 2012; Bharadwaj et al., 2017) or around birth (Nilsson, 2009; Reyes, 2011)) exposure to air pollution on students' scores and do so typically in developed country settings.

While many studies focus on the cumulative or long-run impacts of air pollution on pupils' academic performance, few pay heed to the contemporaneous effect of pollution exposure on student performance. One exception is Ebenstein et al. (2016) who find negative effects of air pollution exposure on the day of high-stake examinations in Israel, with negative long-run effects on individuals' earnings. Another relevant study is Graff-Zivin et al. (2020) who examine the impact of agricultural fires on Chinese students' exam scores. They differentiate between upwind and downwind fires and find the difference be-

tween the two to have a statistically significant effect on exam performance. With regards to Brazil, Bedi et al. (2020) conduct an experiment at the University of São Paulo and suggest that exposure to high levels of $PM_{2.5}$ reduces students' performance in a reasoning test. Thus, we add to this work by investigating the impacts of particulate matter on students' scores in high-stake exams by using a unique geographic and demographically representative data set. Despite the findings of Graff-Zivin et al. (2020) and Bedi et al. (2020) there remains very limited evidence on the impacts of urban air pollution on students' cognitive performance in the most highly polluted cities and regions of the developing world.

This paper examines the causal relationship between outdoor air pollution levels on the day of the nationwide university entry examinations and students' cognitive performance in Brazil, and specifically in Rio de Janeiro and São Paulo, its most industrialised states. More precisely, we use Brazilian data on concentrations of Ozone (O_3) and particulate matter (PM_{10}) and a rich administrative data set on student's scores to examine the impact of air pollution on academic performance at the national examinations *Exame Nacional do Ensino Médio* (ENEM).[2]. This educational data set is combined with air pollution and weather monitoring station data to build a unique panel data for the period 2015 to 2017. We also test the effects of air pollution on the exam performance of male and female students separately since the previous literature has indicated that fetal exposure to pollution can have different long term impacts on males and females (Jayachandran, 2009; Sanders, 2012). Similarly Ebenstein et al. (2016) found boys' exam performance to be more affected by air pollution than that of girls. In addition, we utilise students' socio-economic characteristics to investigate the heterogenous effects of air pollution on students' scores per different economic strata. In line with other heterogeneity analysis, our results suggest that exam takers from poorer socio-economic backgrounds are more impacted by air pollution than their better-off counterparts.

[2] PM_{10} represents particulate matter with a diameter of 10 micrometers or less.

Our analysis contrasts with Ebenstein et al. (2016) who consider the impact of an overall air quality index on the Israeli national examination composite score, and with Graff-Zivin et al. (2020). The latter work could neither establish the direct link between fires and air pollution nor assess the direct effect of air pollutants on students' scores due to lack of data on air pollution for the period of study. Unlike the previous literature, besides using students' fixed effects, we also implement an instrumental variable approach to mitigate endogeneity concerns related to air pollution. Following Deryugina et al. (2019a) and Bondy et al. (2020) we use wind direction as an exogenous shock capable of producing variation in air pollution levels on a particular day and locality. Our assumption relies on the fact that municipalities' daily wind direction is not directly related to students' scores in the national examinations except through its influence on air pollution. We take advantage of the geographic and temporal variation in wind direction in our data set to explore the exogeneity of our instrumental variable, i.e., the variation in daily wind direction in the municipality where the examination took place.

Our focus on Brazil is motivated by the fact that, by global standards, it experiences relatively high levels of air pollution and relatively low levels of academic performance. In terms of pollution, for instance, in 2015 the Environmental Company of São Paulo (*Companhia Ambiental do Estado de São Paulo* (CETESB)) reported 972 occurrences of ozone levels being above WHO safe limits in the metropolitan area of that state.[3] Regarding academic achievements, in the Organisation for Economic Co-operation and Development (OECD)'s Program for International Student Assessment (PISA) Brazil was ranked 39^{th} out of 40 countries in 2019, revealing the relative underperformance of Brazilian students in comparison with their international counterparts, including most of Brazil's Latin American neighbours. Furthermore, OECD research by Zapata et al. (2015) shows that only approximately 15% of 25-34 year-olds in Brazil completed tertiary education in

[3]The information can be found in the CETESB website: https://cetesb.sp.gov.br/.

2014, compared to the OECD average of 41%.

Our findings suggest that an increase of 10μ/m3 of PM_{10} in the day of the examination decreases students' scores by 6.1 points (8%SD). Male students seem to be more affected than female, with a fall of 8.6 points (11%SD) for boys and of 4.3 points (6%SD) for girls. Furthermore, we find students from poorer households to be more affected by an additional 10μ/m3 of PM_{10}, with scores decreasing by 9%SD for those in the bottom 5% of the income distribution. In contrast, individuals from the top 5% are shown not to be significantly influenced by an increase in pollution levels during the tests. We also provide evidence that weaker students tend to perform even worse than top students, with the former suffering a drop of 6%SD and the latter not significantly affected by air quality deterioration. Finally, we run placebo tests using levels of pollution of one day after the exams, with results that reinforce our main findings that it is indeed the poor air quality in the day of the exams that drives scores to descend. That is, the day after the tests has no significant impact on students' cognitive performance. We therefore believe our study provides policymakers with important evidence of a contemporaneous link between air pollution and exam performance.

The remainder of this chapter is organized as follows. In Section 3.2 we discuss previous literature, while in Section 3.3 we describe our data. We present our empirical strategy in Section 3.4. In Section 3.5, we present our results and conduct sensitivity analysis. In Section 3.6 we present our conclusions.

3.2 Literature on air pollution and cognitive performance

An extensive literature has investigated the impacts of air pollution on human health, human capital formation, productivity and the economic growth of countries. In this

section we summarise the literature on the adverse effects of air pollution exposure on cognitive performance[4]. Researchers have demonstrated that the learning process may be affected by air pollution via the health channel due to its harmful effects on the respiratory system. Although it is known that most pollutants affect the respiratory and cardiovascular systems, as well as damaging the circulatory system by obstructing the flow of oxygen to the brain little is known about the channels through which air pollution harms humans health (Seaton et al., 1995; Currie et al., 2009). PM_{10} small particulates enter in the lung, worsening respiratory and cardiovascular problems such as asthma (Dockery et al., 1993; Hansen and Selte, 2000; Currie et al., 2009; Ling and van Eeden, 2009). Ground-level O_3 can also harm lung function and the respiratory system. Studies have shown that exposure to O_3, and to the pollutants it stems from, is related to premature death, asthma, bronchitis, heart attack, and other cardiopulmonary illnesses (Weinhold, 2008; Currie et al., 2009; Neidell, 2009; Developers, 2016; Wikipedia, 2019). Moreover, other papers have demonstrated that even at low levels of O_3 ($40nmol/mol$), more susceptible individuals are more likely to be harmed when exposed to this pollutant (Bell et al., 2006; Weinhold, 2008). There is also compelling evidence that air pollution, through the health channel, increases school absenteeism (Currie et al., 2009; Ransom and Pope, 2013; Chen et al., 2018; Liu and Salvo, 2018), negatively affects school pupils scores (Zweig et al., 2009; Reyes, 2011; Bharadwaj et al., 2017), and decreases low and high-skilled workers productivity (Chang et al., 2016; Heyes et al., 2016; Archsmith et al., 2018; Chang et al., 2019; Kahn and Li, 2019).

School absenteeism

Currie et al. (2009) examine the effect of air pollution on school absences of elementary and middle school students in Texas. They use a panel data from 1996 to 2001 with schools within 10 miles from the air pollution monitors and differences-in-differences-in-

[4]This literature review is not exhaustive. For a thorough review, please, refer to Currie et al. (2014).

differences to tackle the endogeneity stemming from parents address decisions and find that carbon monoxide (CO) significantly increases absences, but PM_{10} presents mixed effects.

Taking advantage of a natural experiment due to the temporary closure of large local steel mill in Utah Valley in the US, Ransom and Pope (2013) uses the mill closure as an instrument for air pollution and find that that particulate pollution has a strong impact on pupils' absence, while CO does not. Although their Ordinary Least Squares (OLS) results underestimate the effect of particulates on school absenteeism, the IV estimates of the effect of CO are, contrary to expectations, negative effects.

Looking at the impact of air pollution on Chinese students' health and school attendance, Chen et al. (2018) use temperature inversion as an exogenous shock for daily air pollution as an instrumental variable. The study matches the school level data with readings from air pollution monitors within a 40 km radius and find a positive relationship between O_3, $PM_{2.5}$, Sulfur Dioxide (SO_2), CO, and AQI and school absenteeism through the health channel.

In a similar study, Liu and Salvo (2018) build a student level panel to investigate the effects of $PM_{2.5}$ on students' absences from international schools in North China. They find a moderate absence rate stemming from the exposure to $PM_{2.5}$, with United States (US)', Canadian and European students more susceptible to skip school days due to air pollution than their Chinese counterparts. This heterogeneity may be due to the long-run adaptation actions by the Chinese parents who are more used to high levels of pollution in comparison to the aforementioned countries.

School examinations scores and human capital formation

Demonstrating the mechanisms through which air pollution affects students' academic performance is a challenging task as tests scores unveil cumulative knowledge. Therefore researchers need to understand the effects of cumulative exposure to air pollution on students' scores (Currie et al., 2014). One of the mechanisms may be through the impact of air pollution on school absence and, ultimately, academic performance. For example, Zweig et al. (2009) investigate the effects of outdoor air pollution levels on standardised state test scores of California public school pupils using the variation in community pollution over time. Using school fixed effects models they find that a 10% decrease in outdoor PM_{10}, $PM_{2.5}$, or Nitrogen Dioxide (NO_2) would raise Mathematics tests scores by 0.15%, 0.34%, and 0.18%, respectively, while a 10% decrease in outdoor $PM_{2.5}$ increases reading scores by 0.21%.

Using Swedish administrative data, Almond and Edlund (2009) investigate the impact of early life exposure to radioactive particles that came from the Chernobyl nuclear disaster on students' scores. They find that the cohorts who were exposed to radiation in utero had their overall scores reduced by roughly 2.5 percentage points in comparison with the pupils not exposed to the pollutant. Also in Sweden, Nilsson (2009) takes advantage of the leaded gasoline ban during the 1970s to assess the effects of lead exposure around birth on pupils' academic performance. The author finds that the reduction in ambient lead is associated with an increase in high school graduation rates by 2.3%, with higher effects for less well-off students.

Reyes (2011) looks at the long-run effects of lead pollution exposure in early childhood in the 1990s on pupils' tests scores in the 2000s. The author uses individual-level data on childhood lead levels and test scores in Massachusetts, controlling for schools and community characteristics and find that high levels of blood lead in early childhood adversely impact standardised test scores. Sanders (2012) investigates the long-run effects of fetal exposure to Total Suspended Particulates (TSPs) on High School students' scores

in Texas. The author takes advantage of an industrial recession in the early 1980s as an exogenous shock to TSPs to establish the negative causal impacts of this pollutant on the students' scores. He also show evidence that boys present worse academic performance than girls due to the prenatal exposure to TSPs.

Examining the impacts of fetal exposure to air pollution on fourth-grade test scores in Santiago, Chile, Bharadwaj et al. (2017) find a strong negative effect from fetal exposure to CO and correlated pollutants, such as PM_{10}, on Mathematics and language skills. Their identification strategy relies on a plethora of fixed effects to control for parents characteristics and siblings fixed effects, since their data on maternal characteristics misses some crucial unobservable family features that may be related to test outcomes and pollution exposure. Additionally, the work includes controls for air quality alerts to account for time-varying avoidance behaviour the parents may engage in to reduce their exposure to air pollution during periods of higher pollutants emissions. The authors find that the 50% reduction in CO in Santiago between 1990 and 2005 increased lifetime earnings by roughly US$100 million per birth cohort.

Most recently two works have addressed the impacts of contemporaneous air pollution on cognitive performance. Firstly, Ebenstein et al. (2016) investigate the contemporaneous effects of air pollution on students' performance at high stakes examinations in Israel. Controlling for a plethora of fixed effects on students' and parents' characteristics, and weather variables, the authors find that the students receive lower scores when exposed to higher levels of air pollution in the exact day of the exams. Secondly, Bedi et al. (2020) have studied the effects of poor air quality on a series of cognitive test in São Paulo. Their findings suggest that the exposure to high levels of $PM_{2.5}$ causes impaired performance on some of the cognitive abilities tested in their experiment.

Low and high-skilled workers' productivity

As a result of the impacts of air pollution on the learning process through the health channel, there is large literature investigating the effects of air pollution on low and high-skilled workers' productivity. For instance, Chang et al. (2016) look at the impact of outdoor $PM_{2.5}$ on the productivity of indoor pear-packing workers. They find that the pollutant negatively affects worker productivity, whereas O_3 has little or no effect. In the same token, Chang et al. (2019) examine the effect of air pollution on call center workers' productivity in China. Using individual level data and controlling for possible meteorological confounders, the authors find that a 10-unit increase in the Air Pollution Index (API) reduces the number of daily calls (per worker) by 0.35%.

Heyes et al. (2016) investigate the effects of $PM_{2.5}$ on high-skilled workers of the New York Stock Exchange. The authors report that a one standard deviation increases in ambient $PM_{2.5}$ reduces same-day S&P 500 returns by 11.9%, suggesting that air pollution increases the operators' risk aversion via pollution-induced changes in mood or cognitive function. Archsmith et al. (2018) also look at the air pollution impact on work performance of quality-focused baseball umpires. The authors take advantage of the fact that these professionals are quasi-randomly assigned across time and matches venues, to estimate the effects of CO and $PM_{2.5}$ on the umpire's decision making. Using fixed effects and a variety of weather controls, they find that a 1 ppm increase in 3-hour CO increases by 11.5% the likelihood of umpires to make incorrect calls and a 10 μ/m3 increase in 12-hour $PM_{2.5}$ causes a 2.6% increase.

Using individual level data from China, Kahn and Li (2019) show evidence that judges (high-skilled workers) are less productive on more polluted days. After controlling for a variety of fixed effects and weather conditions, the authors find that the duration of lawsuits rises by 19.8% due to an increase of 1% in the indoor level of $PM_{2.5}$ the judges are exposed to. This study also demonstrates that the public sector's productivity is more

impacted by air pollution than the private sector's.

3.3 Data and Summary Statistics

Our data set on daily air quality, weather conditions and students' ENEM scores in São Paulo and Rio de Janeiro between 2015 and 2017 is collected from three main sources as detailed below.

3.3.1 Air pollution data

Air pollution data from São Paulo and Rio de Janeiro is collected from their respective environmental state agencies: the Environmental Company of São Paulo (*Companhia Ambiental do Estado de São Paulo*, CETESB) and the State Institute for the Environment (*Instituto Estadual do Meio Ambiente* (Inea)). Both agencies are responsible for monitoring the air quality of the two states, using ground-level automatic and non-automatic monitoring stations.[5] São Paulo has 63 automatic monitoring stations and 26 non-automatic stations, while Rio de Janeiro has 58 automatic and 116 non-automatic stations (see figure B3). Inea's and CETESB's websites provide the addresses, latitude and longitude of their monitoring stations, which we used to merge this data set with the national examination scores, by municipality centroid location.

Unfortunately, for confidentiality reasons, the Ministry of Education in Brazil does not provide the exact address where each student took their examinations, only the municipality. Hence, we are not able to identify the closest monitoring station to the exact exam location. Nevertheless, following Graff-Zivin et al. (2020) we proxy each exam venue's location using the centroid of the municipality in which the test took place. We then identify up to 3 pollution monitoring stations within 10km of the municipality centroid

[5]The automatic stations are directly linked to a central computer that registers the pollution concentrations. In the non-automatic stations, the samples are manually collected in the site to be analysed in the laboratories of CETESB and Inea.

and average hourly readings across those to form an average daily measure of pollution.[6] We then calculate the Inverted Weighted Distance (IWD) for PM_{10} and O_3 (in microgram per cubic meter, μ/m3) following Currie and Neidell (2005); Bondy et al. (2020) and Sager (2019).

3.3.2 Weather data

To control for atmospheric climate we collected data on weather conditions from CETESB and Inea monitoring stations. This provides us with measures of temperature (in degrees Celsius, ^0C), relative humidity (in percentage, %), wind speed (in meters per second, m/s) and wind direction (in degrees)[7]. The latter is used as an instrumental variable for PM_{10}. CETESB provides wind direction in cardinal mode, thus, to use it as a numerical variable, we transform it into degree direction. Moreover, as wind direction is an angle (360^0 circle) and covaries with wind speed (i.e., there are two components that need to be averaged together) to calculate its daily average we need to split out the East/West vector and the North/South vector, and then recombine them to find the speed and direction (Webmet, 2019; Itron, 2019). [8] The remaining climate measures are also registered on an hourly basis, with the daily means calculated and assigned to the examinations municipalities using the IWD in a similar manner to which the pollution measure is built (Currie and Neidell, 2005; Sager, 2019; Bondy et al., 2020). Table 3.1 presents the summary statistics for weather conditions per exam day.

[6] To be clear, if there are three monitors within 10km we average across those, if there are two we average across those and if there is only one then we use the daily average of that one monitor. If a municipality has no monitoring stations within 10km of its centroid then data for that municipality is dropped from our analysis. In our sensitivity analysis we increase the 10km limit to 50km.

[7] Unfortunately, we do not include rainfall as most stations do not present complete data for this variable for the period analysed in this research.

[8] As in Deryugina et al. (2019a), wind direction refers to the direction the wind is blowing from.

Table 3.1: Summary statistics air pollution, weather and exams scores

Variable	All	Girls	Boys
Exam-level data			
Total Score	526	516	537
(0 - 1000 points)	(76.2)	(71.3)	(78)
Sciences	502	493	515
(0 - 1000 points)	(78.9)	(76.4)	(80.2)
Humanities	558	550	569
(0 - 1000 points)	(77.3)	(76.2)	77.5)
Languages	532	532	534
(0 - 1000 points)	(68.4)	(67.1)	(70)
Mathematics	512	494	535
(0 - 1000 points)	(116)	(106)	(125)
PM_{10} ($\mu g/m^3$)	17.5	17.5	17.5
	(6.11)	(6.13)	(6.08)
O_3 ($\mu g/m^3$)	41.5	41.5	41.6
	(12.3)	(12.3)	(12.1)
Temperature (^0C)	20.6	20.6	20.6
	(3.2)	(3.25)	(3.24)
Humidity (%)	81.4	81.4	81.5
	(13.6)	(13.6)	(13.6)
Wind speed (m/s)	2.25	2.25	2.25
	(1.25)	(1.24)	(1.25)
Observations	2,931,368	1,671,910	1,259,458
N Municipality	36	36	36

Note: Table reports statistics for the estimation sample. Standard deviation between parenthesis. N municipalities attends for the total number of municipalities where the exams took place with weather and air pollution measures available for the three years in our main sample.

3.3.3 University entrance examination scores and other ENEM data

The subject specific ENEM scores are collected from the National Institute for Educational Research *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP), an agency from the Brazilian Ministry of Education (INEP, 2017b). This is a rich administrative data set with detailed information on the students' scores per exam subject at the student and school level, and demographic information, such as household income, parental level of education, age, race, gender and special needs (Table 3.1 summarises this information). We build a repeated cross-sectional dataset at the student level from 2015 to 2017.

The data for 2015 indicates that exams took place in 261 municipalities: 208 municipalities in the state of São Paulo and 53 in Rio de Janeiro state. In 2016, the dataset indicates 209 municipalities in the state of São Paulo and 51 in Rio de Janeiro state, making a total of 260 municipalities where the ENEM took place. In 2017, the state of São Paulo held examinations in 209 municipalities, and the state of Rio de Janeiro in 52. Note that since we use monitoring stations as far as 10km from the municipality centroid to construct our air pollution and meteorological measures, we end up with only 36 municipalities within our final sample. This reduction in the number of municipalities and hence number of spatial observations of pollution measures in our final sample consists of a trade-off decision that we must take. Had we opted for using stations located, for instance, 50km from the municipality centroid, we would have experienced attenuation bias in our estimates [9]. Nonetheless, the reduction in the number of observations in our final panel does not affect the power of our study since we still end up with a very large sample size.

[9] Table B4 in the appendix presents the number of observations and municipalities for the 5 samples using the different distances.

As for demographic information, the INEP data provide the highest level of education achieved by each student's parents. As such, we build dummies indicating each of the seven educational levels. We create eight dummies for each parent's education: (i) Never studied, (ii) Incomplete 1^{st} elementary if the parent attended school for a while, but did not complete the first elementary, (iii) Incomplete 2^{nd} elementary if the parent attended school for a while, but did not complete the second elementary, (iv) Elementary if the parent only completed the Elementary School, (v) High School if the parent completed the High School, (vi) Undergraduate if the parent completed Undergraduate, (vii) Post-Graduate if the parent completed Post-Graduate level, and (viii) non-declared if the students declared they were unaware of their parents' level of education. (Zweig et al., 2009).

Students' socio-economic background is also likely to be an important predictor of their academic performance, particularly in developing countries where the degree of inequality is much higher than in developed nations. In Brazil it has been found that students' socio-economic characteristics can determine up to 85% of the variation in the scores at the exam (Estadao, 2019). In light of that, in our models that use municipality fixed effects, we control for socio-economic background using parental household income. The original data does not display the income in figures, preventing us from using it as a continuous variable. As a result, we created dummies for each level of parental household income, dropping the missing values. According to the INEP's micro-data information, there are 17 income ranges - monthly earnings from zero minimum wage until more than 20 minimum wages -, thus we created 17 dummies, including 16 in our models.

The ENEM

The ENEM is a non-mandatory and standardised national exam used to evaluate high school students in Brazil and is the second largest in the world after the National Higher Education Entrance Examination in China. Since 2009, the exam has been used by fed-

eral and state universities as their admission test. The exam is also used to rank deprived students to receive points in the federal scholarship *Universidade para Todos Program* (or ProUni), as well as for certification for a high school degree.

The ENEM comprises four groups of multiple choice exams that take place in two days: Humanities, Sciences, Languages, and Mathematics, each of them with scores ranging from 0 to 1,000. The Humanities exam includes History, Geography, Philosophy and Sociology, the Sciences tests engulfs Chemistry, Physics and Biology, and the Languages includes Portuguese Language, Literature, Foreign Language (English or Spanish), Arts, Physical Education, Information and Communication Technologies (INEP, 2017b). In 2015 the examinations occurred on the 24^{th} and the 25^{th} October; in 2016, on the 5^{th} and the 6^{th} November; and in 2017, on the 5^{th} and the 12^{th} November. In each year the exams start at 13h30min and finish at 18h30min on the first day, and start at 13h30min and finish at 19h00min on the second day. In 2015 and 2016 on the first day, students took exams on Humanities and Sciences; and in the second day, Languages and Mathematics. In 2017 the order was slightly different, with students taking tests on Languages and Humanities in the first day, and Sciences and Mathematics in the second day. Although we have hourly data for both air pollution and weather variables, which would enable us to assign the measures of air pollution and weather condition at the exact moment of the exams, we follow the literature on the impact of air pollution on human capital and use the daily average of O_3, PM_{10}, humidity, temperature, wind speed, and wind direction (Ebenstein et al., 2016; Sager, 2019; Bondy et al., 2020).

3.4 Empirical Strategy

We examine the partial correlation between PM_{10} exposure and test scores at the individual-level for the years 2015 to 2017. Nonetheless, we know that there are a few

identification challenges to establish the causal relation between air pollution and students' performance due to the possible presence of omitted correlated elements. For instance, one concern is that students' exam scores may be correlated with our two measures of air pollution due to the fact that students from higher socio-economic strata could live in municipalities with lower levels of air pollution (the parents could choose to live in less polluted areas, which are usually more expensive) (Banzhaf and Walsh, 2008; Currie et al., 2014). Hence, although the candidates do not choose the exam venue, it is determined according to the student's household postcode, i.e., INEP tries to assign the examination venue as close as possible (up to 30 km) to each student's address. Hence, students who live in less polluted municipalities would end up taking exams in cities with cleaner air than other candidates who live in more deprived areas, which may be more polluted (Currie et al., 2014). This possibility of address selection is also highlighted by Neidell (2004) and Neidell (2009), who characterise it as an avoidance behaviour by parents.

A final concern relates to the fact that the level of pollution and weather conditions in the days and venues where the exams are marked may influence the final marks as the examiners could also have their productivity affected by these circumstances. However, all the tests are multiple choice and the grades are given by a machine, which clearly is not affected by air pollution or weather determinants. Thus, to tackle those econometric concerns we apply two specification strategies as detailed below.

3.4.1 Panel Fixed Effects Model

As students take exams in two days, we use students fixed effects and control for the exact day of the examinations. The main model thus takes the following form:

$$S_{imt} = \beta PM_{mt} + \tau TEMP_{mt} + \rho HUMID_{mt} + \psi WINDSPEEDmt + \phi_i + \theta_t + \epsilon_{imt} \tag{3.1}$$

where S_{imt} is the test score of student i at municipality m at time t, PM_{mt} is our measure of air pollution (PM_{10}) (in μ/m3) at municipality m at time t, $TEMP_{mt}$ is the daily average temperature at municipality m at time t in degrees Celsius, $HUMID_{mt}$ is the relative humidity measure at municipality m at time t in percentage[10], $WINDSPEED_{mt}$ is the mean of wind speed measured at municipality m at time t in m/s; ϕ_i is students fixed effects. By including fixed effects and weather controls, we are able to compare students' performance in the exams in the different days with different levels of air pollution given each student's characteristics. Hence, confounding factors stemming from time-invariant differences between students can be eliminated from our list of identification concerns. For example, socio-economic background and parents' level of education are subsumed by the students fixed effects. θ_t controls for day of exam fixed effects. As we know the day of each examination, we are able to add this control and rule out the concern that the tiredness of students in the second day of exams would confound the real effects of air pollution on their performance due to extra exhaustion in the second day of tests. In addition, we can eliminate another concern regarding differences in level of pollution from Saturdays and Sundays (when the exams take place) by including these time-varying controls. Finally, the standard errors are clustered by municipality centroid to control for spatial and serial correlation within each municipality in which examinations took place. and ϵ_{imt} is an idiosyncratic error term.

[10] In our Econometric models, we add linear and quadratic terms in both temperature and relative humidity, and their linear and quadratic interaction.

3.4.2 The instrumental variable approach

After using students fixed effects and controlling for weather conditions, there may still remain concerns on the existence of classical measurement error. For example, the students' actual exposure to outdoor air pollution may differ from the monitoring stations readings, which may bias our estimates downward due to attenuation bias (Moretti and Neidell, 2011; Schlenker and Walker, 2015; Chen et al., 2018). Furthermore, as explained above, as PM_{10} is not randomly assigned, leading to the potential presence of unobserved time-varying effects that we cannot account for. Therefore, we also experiment with an instrumental variable approach to estimate our models. In this regard we follow previous works, as, for example, Anderson (2015); Deryugina et al. (2019a); Bondy et al. (2020), and adopt wind direction as our instrument for air pollution, which can be considered as a natural experiment that results in exogenous shocks to local air pollution in a specific area due to wind shocks (Currie et al., 2014).

To identify the causal relationship between pollution and the students' scores we follow the strategy employed by Deryugina et al. (2019a) and Bondy et al. (2020) to build our instrumental variable. To this end we rely on the assumption that municipalities' daily wind direction is not directly related to the students' scores in the national examinations, except through its influence on air pollution. Therefore, as argued by Deryugina et al. (2019a), we do not need to know the exact source of pollution since using wind direction - which is totally random - to predict air quality eliminates the concern of potential correlation between what generates pollution and other factors affecting student performance. We take advantage of the geographic and temporal variation in wind direction in our data set as the source of exogeneity of our instrumental variable. That is, we use the variation in daily wind direction in the municipality where the examination took place as

an instrument for pollution. Our model can be formally described as:

$$Pollutant_{imt} = \sum_{g \in G} \sum_{b=0}^{2} \beta_b^g 1\left[G_c = g\right] \times WindDir_{mt}^{90b} + \theta_1 temp_{mt} + \theta_2 humidity_{mt} + \theta_3 windspeed_{mt} + \nu_i + \phi_t + \epsilon_{imt} \tag{3.2}$$

$$Score_{imt} = \alpha \widehat{Pollutant}_{mt} + \psi_1 temp_{mt} + \psi_2 humidity_{mt} + \psi_3 windspeed_{mt} + \mu_i + \omega_t + \upsilon_{imt} \tag{3.3}$$

where $Score_{imt}$ is the test score of student i at municipality m at time t. The variable $1\left[G_c = g\right]$ indicates the monitor group g each municipality was classified into out of the total 3 G. The set of variables in $1\left[G_c = g\right] \times WindDir_{mt}^{90b}$ account for our excluded instruments, where each of the g clusters is interacted with $WindDir_{mt}^{90b}$ and equals 1 if the daily average wind direction in the municipality m is in the 90-degree interval [90b,90b+90) and 0 otherwise, with the interval [270,360) omitted (Deryugina et al., 2019a). Equation 4.4 is the first stage in which we expect $\hat{\beta}$ to be significant as an indication that wind direction at municipality m and time t affects the level of pollution at municipality m and time t (relevance assumption).

Following Deryugina et al. (2019a) and Bondy et al. (2020), we use k-means cluster to create 3 groups for all the pollution monitors in our sample. Clustering is largely used to assign pollution monitors into a pre-determined number of groups according to the groups' latitude and longitude. Hence, each group represents a geographic area [11]. To choose the optimal number of regional clusters, we use the k-means cluster algorithm (Makles, 2012). First, we follow the official IBGE regional division of the São Paulo and Rio de Janeiro States to define 11 the number of maximum possible clusters the municipalities in our

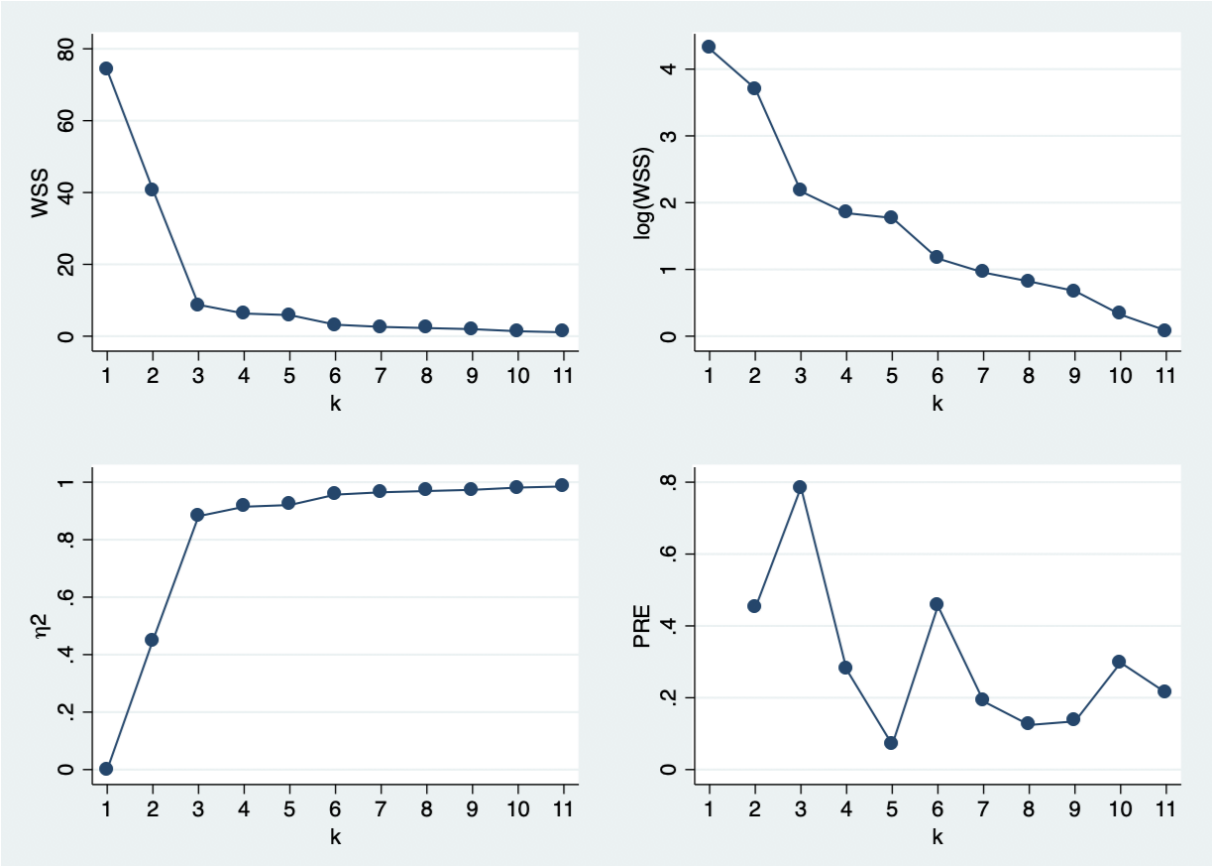
[11] K-means clustering aims to split n observations into k clusters in which each observation belongs to the cluster with the nearest mean. K-means algorithm minimizes within-cluster variances (squared Euclidean distances) and tends to find clusters of comparable spatial extent; i.e., by latitude and longitude in our case.

final sample could belong to. The IBGE adopts several forms to classify the territory into smaller regions, for example macro-regions, micro-regions and meso-regions. We adopted the latter classification in line with CETESB's and Inea's monitoring stations organisation to define the municipalities from our sample into those regional groups. We found 11 regional levels, which we inputed in the k-means algorithm to find the optimal number of clusters per latitude and longitude. Figure 3.1 shows the Within Sum of Squares (WSS), log(WSS), η^2, which is similar to the R^2, and the Proportional Reduction of Error (PRE) coefficient for all k cluster solutions. Accordingly, clustering with k=3 appears to be the optimal solution. More specifically, at k=3 we see a kink in the WSS and log(WSS). η^2 points to a reduction of the WSS by 88% and PRE to a reduction of about 79% compared with the k = 2 solution. Yet, the reduction in WSS is negligible for k > 3 [12].

Note that the number of groups are somewhat (by latitude and longitude) arbitrarily chosen (Deryugina et al., 2019a; Bondy et al., 2020). As a result, β_b^g varies across geographic groups (Deryugina et al., 2019a). It follows that, in using the interaction between the geographic groups with the wind direction dummies, we aim to mitigate the concern that a municipality's pollution monitors may not properly measure the average pollution each student within each municipality is affected by since the pollution station locations are sparse throughout the two states Rio de Janeiro and São Paulo. Furthermore, we also do not know the precise location of the exams venues within each municipality. As such, our identification strategy limits the effect of municipality wind direction on pollution to be the same for all stations within each of the 3 geographic groups, which comprise various municipalities. As explained by Deryugina et al. (2019a) and by Bondy et al. (2020), this confinement shortens the impact of pollution variation from different sources, reducing measurement error in our models. The concern here is that the 36 municipalities in our sample engulfs a large area and wind may carry particulate matter from different

[12] Our results do not leave doubt concerning the k=3 choice. Nonetheless, we k-means response depends on the the initial cluster centres chosen by the researchers. Thus, we repeated the process 50 times and found k=3 to be the optimal answer 83% of the times. Furthermore, for more details on how to calculate the WSS, η^2 and PRE, consult Makles (2012).

Figure 3.1: WSS, log(WSS), η^2, and PRE for all K cluster solutions (Optimal number of clusters



Note: The figure indicates clustering with k = 3 to be the optimal solution. At k = 3, there is a kink in the WSS and log(WSS), respectively. η^2 points to a reduction of the WSS by 88% and PRE to a reduction of about 79% compared with the k = 2 solution. However, the reduction in WSS is negligible for k > 3.

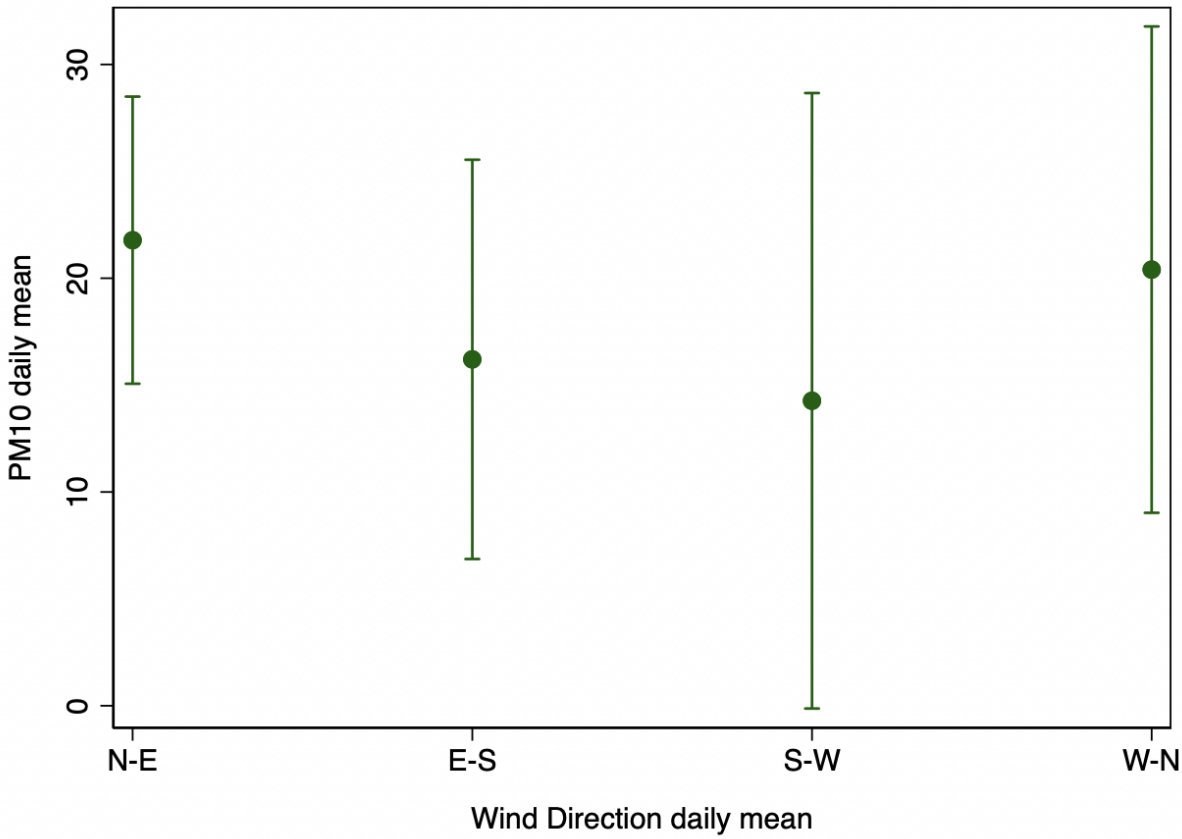
sources [13]. That is, if the station is close to the point source, its measures are likely to be higher than another station located far from the point source, which may cause measurement error in our estimations. On the other hand, nonpoint sources tend to present more similar influence on the monitors in the same group; and hence are more likely to be responsible for the variation on pollution that we expect to capture from the variables $1\left[G_c = g\right] \times WindDir_{mt}^{90b}$ (Deryugina et al., 2019a).

From Figure 3.2 shows the structure of our instruments by depicting the daily average PM_{10} across the 36 municipalities where the exams took place in our final sample by wind

[13]Note that we consider the direction the wind blows from, instead of the direction that the wind blows to as in Anderson (2015).

direction measured by its daily mean. It illustrates the relationship from our first stage suggesting that days with wind blowing from north, northeast and east have on average higher levels of PM_{10} [14].

Figure 3.2: The effect of wind direction on air pollution



Note: Daily average of PM_{10} concentrations for the 36 municipalities in the final sample per direction from where the wind blew during the exams days. 95% confidence intervals. Details on the data described in the text.

Our instrumental variable approach confirms our fixed effects results. Primarily, our valid set of instruments helps us to address potential unobserved factors that could remain in our baseline findings. Suffice it to say that our estimates are roughly the same in both empirical strategies. In addition, by using interactions between wind direction and geographical clusters, we are capable of demonstrating plausibly exogenous variation in air pollution measures across different regions comprising our sample.

[14] Table B1 shows the F test statistics for the strength of our instruments.

As for the validity of our instrument, wind direction fulfills the relevance requirement since it varies along with the concentration of PM_{10} as demonstrated by the F-test statistics in Tables B1 and 3.3. Moreover, with regards to the exogeneity requirement, the main assumption is that wind direction only affects students' scores via air pollution.

Although the previous literature has shown that O_3 and PM_{10} are not highly correlated, we use the daily average of pollution measures of PM_{10} as our main pollutant measure (Zweig et al., 2009; Ebenstein et al., 2016; Sager, 2019). This separate equations approach, i.e., the focus on one particular pollutant, has been used by, for example, Currie (2009) and Arceo et al. (2016). Other works prefer to use a composite measure of air pollution. [15] Nonetheless, we estimate equations with both pollutants jointly as a robustness check, and, as Table 3.6 shows, the estimate signs and magnitudes of O_3 and PM_{10} follow the same pattern as when PM_{10} is estimated in individual model.

For the temperature variables we build six bins of the size 3^0C, with the first bin comprising temperatures between 16^0C and 19^0C, and the last bin for temperature higher between 28^0C and 31^0C. This approach allows temperature to have a non-linearity impact. In all specifications, we use linear and quadratic terms for temperature (or bins) and humidity and their interactions (Ebenstein et al., 2016).

3.5 Empirical Results

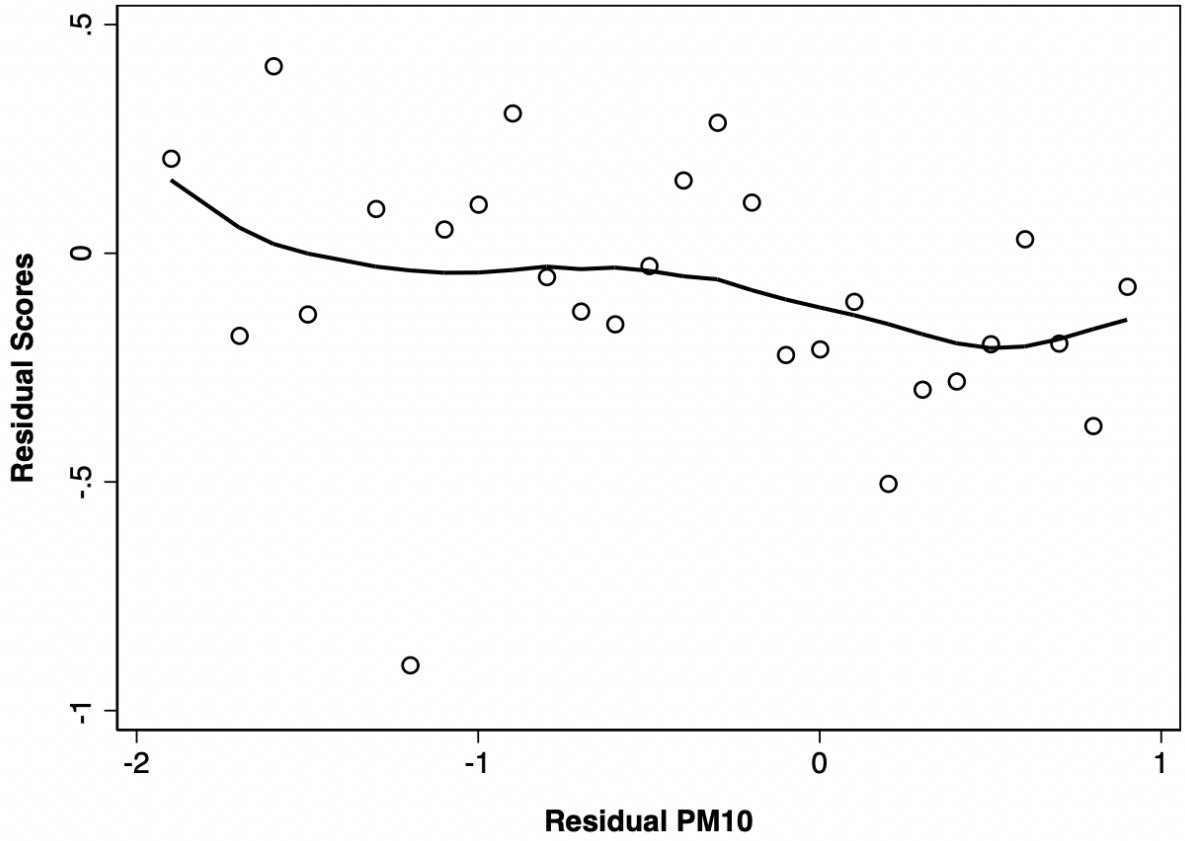
3.5.1 Main Results

Table 3.2 provides our estimation results. For completeness, we start by using municipality fixed effects. From Column 1 we can observe that an additional $10\mu g/m^3$ of

[15] For instance, Ebenstein et al. (2016), Chen et al. (2018) and Bondy et al. (2020) use the AQI.

PM_{10} corresponds to a reduction in students scores by 1%, although not significantly so. This result does not change after we include students' controls such as gender, race and mother's level of education. On the other hand, when it comes to utilising students fixed effects, the results in column 5 suggest that an increase of $10\mu g/m^3$ of PM_{10} on the day of examinations leads to a drop of 4% of a SD, though not significant when we do not control for the day of the exam. After controlling for day fixed effects, we notice, in column 7 - our preferred specification -, a highly significant reduction in students' scores of 8% of a SD for every $10\mu g/m^3$ increase in the same pollutant concentration.

Figure 3.3: Residual PM_{10} and Residual Scores Scatterplot (student FE)



Note: Plot of residual PM_{10} and residual Scores (SD) estimated by lowess bandsmoother. Residuals from a regression on students fixed effects and weather controls.

Figure 3.3 adds to our results in Table 3.2 by depicting that the negative relation between residual PM_{10} and students' scores is evident even when solely controlling for variation within students' idiosyncrasies. The graph is generated using the residuals from a regression of PM_{10} on students fixed effects and weather controls, showing that, on av-

erage, a student exposed to higher level of air pollution presents lower scores than another test taker exposed to better air quality during the exams.

Turning to our IV approach, the first stage's F-test statistic informs that wind direction is indeed a strong predictor of air pollution, hence, a strong instrument. Table B1 complements these results on the strength of our instruments. Our findings in Table 3.3 suggest that an additional $10\mu g/m^3$ of PM_{10} in the days of examinations caused the students' marks to drop by 8% (column 3). If we compare this with the baseline fixed effects results presented in Columns 1 and 2, we notice that the results remain pretty much the same even thought there is a slight reduction in the sample size.

To explore possible non-linearity between air pollution and students' cognitive performance, we present results from a complementary specification using a dummy variable for air pollution that equals 1 if the level of PM_{10} is above $20\mu g/m^3$ - WHO's acceptable threshold for this pollutant. Table 3.2 Columns 7 shows the estimates for our preferred model using linear pollution. One can note that the results are also significant with an increase of $10\mu g/m^3$ of PM_{10} causing a reduction of 11%SD in students' scores. When it comes to non-linear PM_{10} for the instrumental variable specification, Column 4 in Table 3.3 displays the same pattern of results as the one using linear pollution measure (column 3). That is, an additional $10\mu g/m^3$ on the day of the examination reduces students' scores by 13% SD. Also, regarding the first stage for the non-linear approach, the F test statistics still reveals that wind direction is a strong instrument for air pollution. These pieces of evidence suggest that cleaner air may lead to better educational outcomes and economic performance accordingly.

3.5.2 Heterogeneity across sub-populations

In this section, we analyse the heterogeneous effect of the treatment (air pollution) for our preferred specification reported in columns 7 and 8 from Table 3.2: (i) per gender;

(ii) across the distribution of scores per municipality; and (iii) in terms of different levels of income. Through this heterogeneity inspection, we aim to unveil how different sub-populations have their cognitive performance affected by the exposure to air pollution during high-stakes examinations.

With regards to gender differentials, our findings in Table 3.4 suggest that boys are more negatively affected by poor air quality during exams than girls. To put this into perspective, an increase of $10\mu g/m^3$ of PM_{10} causes males' scores to drop by 11%SD, whereas girls suffer a reduction of 6%SD. This heterogeneity is observed even when using a non-linear pollution variable; the results persist with boys' scores being reduced by 15%SD and girls' by 8%SD, meaning that boys who passed the examinations in days with air pollution levels above the WHO's acceptable threshold face a reduction in scores by 7pp deeper than girls (columns 6 and 7). This finding is consistent with previous studies including Jayachandran (2009); Sanders (2012), and Ebenstein et al. (2016). Although the role of gender in the effects of air pollution on respiratory health remains unclear, some epidemiology studies point out that males are more affected than females. For example, Abbey et al. (1998) link PM_{10} exposure to reduced lung function among males but not females, Galizia and Kinney (1999) find similar results regarding the exposure to O_3, and Chen et al. (2017b) identify that contemporaneous and cumulative exposure to air pollution has a stronger negative effect for men than for women. Similarly, Ebenstein et al. (2016) find the effect of pollution on exam performance to differ by gender and cite the increased incidence of asthma amongst boys as a possible cause.

As per the second heterogeneity analysis, our goal is to detect whether strong or middling students are differently affected by poor air quality. To this end we split our final sample into subsets per low and high scores within each municipality. One can note from Panel A Table 3.5 that students below the average have their scores 1pp more impacted than pupils that rank above the average; i.e., stronger students exposed to an additional

$10\mu g/m^3$ of PM_{10} during the exams have their performance reduced by 6%SD (Column 3), whilst weaker students are more negatively affected presenting a drop of 7% SD (Column 2). When we compare to the baseline estimation (column 1) using the whole sample, the average reduction is 8% SD. Another comparison is made between the top 5% scores and bottom 5%. Column 4 shows that students who rank very low in the exam suffer a reduction of 5% on their marks, whilst students on the top of the distribution seem not to be significantly affected by an increase of pollution levels during examinations (column 5). Finally, columns 6 to 10 bring results from the same exercises utilising the non-linear pollution variable. These findings suggest that students above the median and who took the exams in days with PM_{10} above the WHO'S threshold are more negatively impacted, with scores dropping by 12%SD, while students in the top 5% of the ranking are again not significantly affected by taking exams in days with pollution levels higher than $20\mu g/m^3$.

In Panel B Table 3.5, we slice our sample to look into how students from different socio-economic strata may be disproportionately affected by air pollution during the tests (Hsiang et al., 2019). For example, Neidell (2004) finds that less well-off households are more impacted by air pollution given their reduced capability of compensating behaviour. Similarly to the scores' distribution analysis, students from poorer backgrounds seem to be more affected by bad air quality. It may also be due to the fact that weaker students and low income are correlated Ebenstein et al. (2016). We therefore sub-sampled the individuals according to their household incomes and compared them using the mean of the municipality average income. From Panel B Table 3.5, one can notice that students below the median or below the bottom 5% face a reduction of 9%SD in their scores for every extra $10\mu g/m^3$ in the day of the tests (Columns 2 and 4), whereas wealthier students are not significantly impacted by a deterioration in air quality.

Table 3.2: The Effects of PM_{10} on Students' Scores (Main Specification with flexible weather controls)

	Municipality Fixed Effects				Students Fixed Effects			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PM_{10}	-0.01		-0.01		-0.04		-0.08***	
$(10\mu g/m^3)$	(0.03)		(0.05)		(0.03)		(0.03)	
1 if $PM_{10} > 20$		-0.04		-0.04		-0.09		-0.11***
		(0.03)		(0.03)		(0.07)		(0.04)
Students FE	N	N	N	N	Y	Y	Y	Y
Students Controls	N	N	Y	Y	N	N	N	N
Day FE	Y	Y	Y	Y	N	N	Y	Y
Weather controls	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,944,911	2,944,911	2,944,911	2,944,911	2,931,368	2,931,368	2,931,368	2,931,368
R-squared	0.004	0.004	0.17	0.17	0.05	0.05	0.07	0.07

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students controls includes gender, race and mother's level of education (columns 2 and 4), municipalities FE (columns 1 to 4) students FE, day FE (columns 5 and 6: there is no day FE), temperature in bins, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Table 3.3: Particulate Matters's Impact on Scores: Models with Instrumental Variable Wind Direction

Variables	(1) Baseline	(2) Baseline	(3) With Instruments	(4) With Instruments
PM_{10} $(10\mu g/m^3)$	-0.08*** (0.03)		-0.08*** (0.01)	
1 if $PM_{10} > 20$		-0.11*** (0.04)		-0.13*** (0.01)
F First Stage			55.4	71
OBS	2,931,368	2,931,368	2,405,938	2,405,938

Note: See table 3.2 for full specification. Columns 1 and 2 identical to columns 3 and 4 from Table 3.2. Flexible weather controls; students and day FE. Standard errors clustered by municipality and bootstrapped with 500 repetitions.*** p<0.01, ** p<0.05, * p<0.1

Table 3.4: Heterogeneity in Particulate Matter's Impact on Students' Scores (in SD) per gender (with bins)

Variables	(1) All	(2) Girls	(3) Boys	(4) All	(5) Girls	(6) Boys
PM_{10} $(10\mu g/m^3)$	-0.08*** (0.03)	-0.06*** (0.03)	-0.11*** (0.03)			
1 if $PM_{10} > 20$				-0.11*** (0.04)	-0.08** (0.03)	-0.15*** (0.05)
Observations	2,931,368	1,671,910	1,259,458	2,931,368	1,671,910	1,259,458
R-squared	0.07	0.09	0.08	0.07	0.09	0.08

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE, temperature in bins, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Table 3.5: Heterogeneity in Particulate Matter's Impact on students' scores (in SD) across sub-populations (bins)

Variables	(1) Baseline	(2) Below 50%	(3) Above 50%	(4) Below 5%	(5) Above 95%	(6) Baseline	(7) Below 50%	(8) Above median	(9) Below 5%	(10) Above 95%
					Panel A: Low and high scores					
PM_{10}	-0.08***	-0.07***	-0.06**	-0.05***	-0.03					
$(10\mu g/m^3)$	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)					
1 if $PM_{10} > 20$						-0.11***	-0.07*	-0.09***	-0.05**	-0.04
						(0.04)	(0.04)	(0.03)	(0.02)	(0.03)
					Panel B: Low and high socioeconomic status					
	Baseline	Below 50%	Above 50%	Below 5%	Above 95%	Baseline	Below 50%	Above median	Below 5%	Above 95%
PM_{10}	-0.08***	-0.09***	-0.08**	-0.09***	-0.07					
$(10\mu g/m^3)$	(0.03)	(0.02)	(0.04)	(0.02)	(0.06)					
1 if $PM_{10} > 20$						-0.11***	-0.10*	-0.12***	-0.09**	-0.10**
						(0.04)	(0.05)	(0.04)	(0.04)	(0.05)

Notes: The dependent variable in all regressions is students' scores (in SD). Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE, temperature in bins, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

3.5.3 Sensitivity Analysis

To assess the robustness of our results, we experiment with a number of alternative specifications. So far we have reported results in which we include only one pollutant at a time in our estimations. Given the generally high correlation between many common local air pollutants there continues to be debate around whether or not different pollutants have independent health effects (see for example Lipfert et al., 2009; Katanoda et al., 2011; Cao et al., 2011; Gan et al., 2011; Chen et al., 2012). In our sample, O_3 and PM_{10} have a pairwise correlation of 0.10, significant at 5%. Nevertheless, Table 3.6 provides the results from estimations in which both pollutants are included together. In line with Bedi et al. (2020)'s work on cognitive performance in São Paulo, we see that PM_{10} continues to have the same impact on test takers (column 3), whilst O_3 does not present a significant effect on the latter. These are expected results since PM_{10} penetrates indoors easier than O_3, hence one would indeed expect lower or none impact of the latter pollutant in a co-pollutant estimation.

In a second exercise we include a one day lead in our main specification with the linear form of PM_{10} to capture common situations for example political events, weather shocks. We use data on air pollution from one day ahead of the actual date but which contains common features as the true measure. Table 3.7 shows that air pollution levels in the day after the exams do not significantly affect students' performance. Accordingly, an additional $10\mu g/m^3$ in the day of examinations is still responsible to a decrease of 5%SD in students' scores, while PM_{10} measures from the following day have null impact on pupils' marks. These results support our main findings not be driven by other factors that started before the test day and correlated with pollution, but solely by exposure to poor air quality on the day of the test.

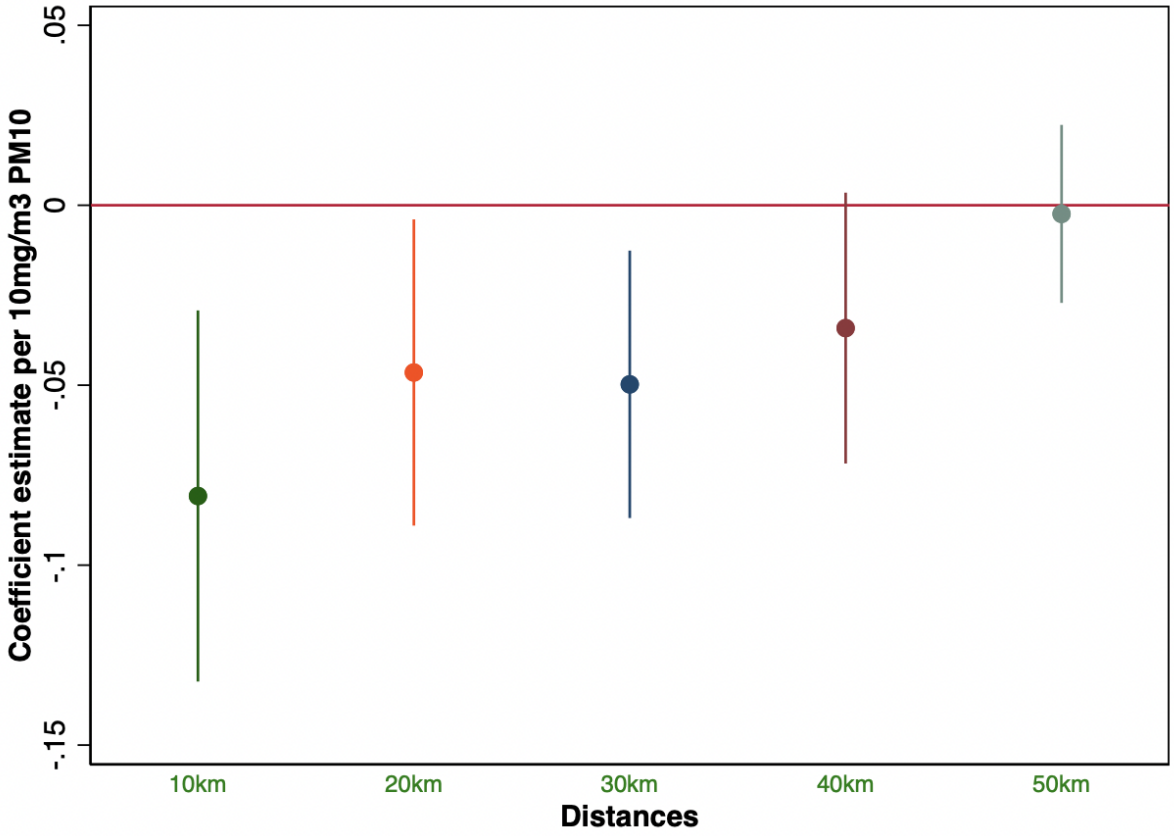
We next conduct a third exercise to check the robustness of our results. As we do not know the precise location where the students took their exams, we have so far used

the municipality centroid as a proxy for the exam venues and measured pollution using the average of up to 3 pollution monitoring stations within 10km of that centroid. However, if a school is located towards the periphery of a municipality then the monitoring stations within 10km of the centroid may not be the best measure of pollution levels at that school's location. Furthermore, we are omitting monitoring stations from our analysis which may be closer to the school's location but further away from 10km. To see whether this is affecting our results we here extend the 10km limit to 50km and plot the results in figure 3.4. We can see that PM_{10} continues to negatively affect boys' scores but the impacts fade and the standard errors inflate insofar as the distance rises, with the coefficients corresponding to the 50km estimation not statistically significant. This would seem to suggest that the 10km cut-off more accurately captures pollution concentrations at exam venues and that the pollutants' levels decay insofar as they travel [16].

Finally, to demonstrate that our main specifications with binned weather controls are not sensitive to different forms of meteorological controlling, Tables B3 and B2 present our main specification results without controlling for weather and with less flexible (with linear; and with squares and interactions controls). The results remain roughly the same, supporting our main estimates not be driven by unobserved weather conditions correlated with wind and students' performance, but solely by the effects of air pollution.

[16] Similar results can be found, for example, in Lai et al. (2018)

Figure 3.4: Cognitive impact of PM_{10} per distance from the station to the municipality centroid



Note: This figure reports the cognitive impact of air pollution within different distances. 10km is the baseline model. All the specifications follow the main model with students FE, day controls and flexible weather condition as in column 7 from Table 3.2. We note an increase of standard errors suggesting that the further the stations the less precise and significant the estimates.

Table 3.6: The Effects of PM_{10} on Students' Scores in SD (with co-pollutant (bins))

Variables	(1) Main model	(2) Main model	(3) With co-pollutant	(4) With co-pollutant
PM_{10}	-0.08***		-0.08***	
$(10\mu g/m^3)$	(0.03)		(0.03)	
1 if $PM_{10} > 20$		-0.10***		-0.05
		(0.04)		(0.04)
O_3			0.01	0.01
$(10\mu g/m^3)$			(0.02)	(0.02)
Observations	2,931,368	2,931,368	2,931,368	2,931,368
R-squared	0.07	0.07	0.07	0.07

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE, temperature in bins, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} and O_3 accounts for their linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Table 3.7: The Effects of PM_{10} on Students' Scores in SD (Baseline Model + 1 day lead)(bins)

Variables	(1) Baseline	(2) Baseline + lead	(3) Lead
PM_{10}	-0.08***	-0.05**	
	(0.03)	(0.02)	
PM_{10} (Lead)		-0.00	0.02
		(0.02)	(0.02)
Observations	2,931,368	2,172,680	2,172,680
R-squared	0.07	0.08	0.08

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Included students FE, day FE, flexible weather controls. PM_{10} is linear in $10\mu g/m^3$.

3.6 Conclusion

In this paper, we take advantage of the individual-level panel data structure of our data set to estimate the causal effects of contemporaneous exposure to air pollution on students' performance in university entrance exams from 2015 to 2017. To identify this relationship we use students fixed effects as our main specification strategy, and, as robustness checks, we utilise a wide range of control variables, such as individual characteristics, parents' level of education, household income, weather controls (temperature, humidity and wind speed), and day and municipality fixed effects. To address possible bias in our estimations stemming from classical measurement error and omitted variable bias, besides individuals fixed effects, we also perform an instrumental variable approach where we rely on the temporal and geographical variation in wind direction to build our instrument. Our main assumption is that this variation is exogenous since it does not directly affect the students' scores, except from its effect on air pollution.

We find that an increase of 10μ/m3 of PM_{10} in the day of the examination decreases the students' scores by around 6 points (8%SD). On top of that, when we study the response from different sub-populations within our sample to same levels of air quality, we find students from the bottom 5% of the economic ladder to be more responsive to a rise of 10μ/m3 of that pollutant, with a drop of 9%SD , whereas the 5% better-off exam takers present no significantly reaction to the exposure to PM_{10} in the day of the exams. Our results consistently suggest the harmful effect of air pollution on exam performance. Even when including a more flexible measure of our treatment; i.e., utilising a dummy variable to account for the days in which PM_{10} exceeded the WHO's acceptable threshold, our findings still point to the negative effects of air pollution on cognitive performance during examinations. Finally, we run several sensitivity checks and falsification tests, which suggest that our results are not driven by spurious correlations.

Two broad findings emerge from our results. Firstly, the effect of air pollution on

exam performance appears to differ for males and females, with the performance of the former more adversely affected. This finding is consistent with previous studies including Jayachandran (2009); Sanders (2012) and Ebenstein et al. (2016). Although the role of gender in the effects of air pollution on respiratory health remains unclear, some epidemiology studies point out that males are more affected than females. For example, Abbey et al. (1998) link PM_{10} exposure to reduced lung function among males but not females; and Galizia and Kinney (1999) find similar results regarding the exposure to O_3. Similarly, Ebenstein et al. (2016) find the effect of pollution on exam performance to differ by gender and cite the increased incidence of asthma amongst boys as a possible cause. Secondly, our results suggest that the effect of air pollution on exam performance differs by socio-economic background, with poorer students scoring even worse than wealthier treated exam takers. These findings are also in accordance with the literature, (see for example Neidell (2004) and Ebenstein et al. (2016), as there might exist a correlation between low parental income - and hence an impossibility of compensating behaviour - and pertaining to the bottom of the score distribution. All in all, our findings seem to add plausible evidence suggesting that cognitive performance may be hindered by bad air quality.

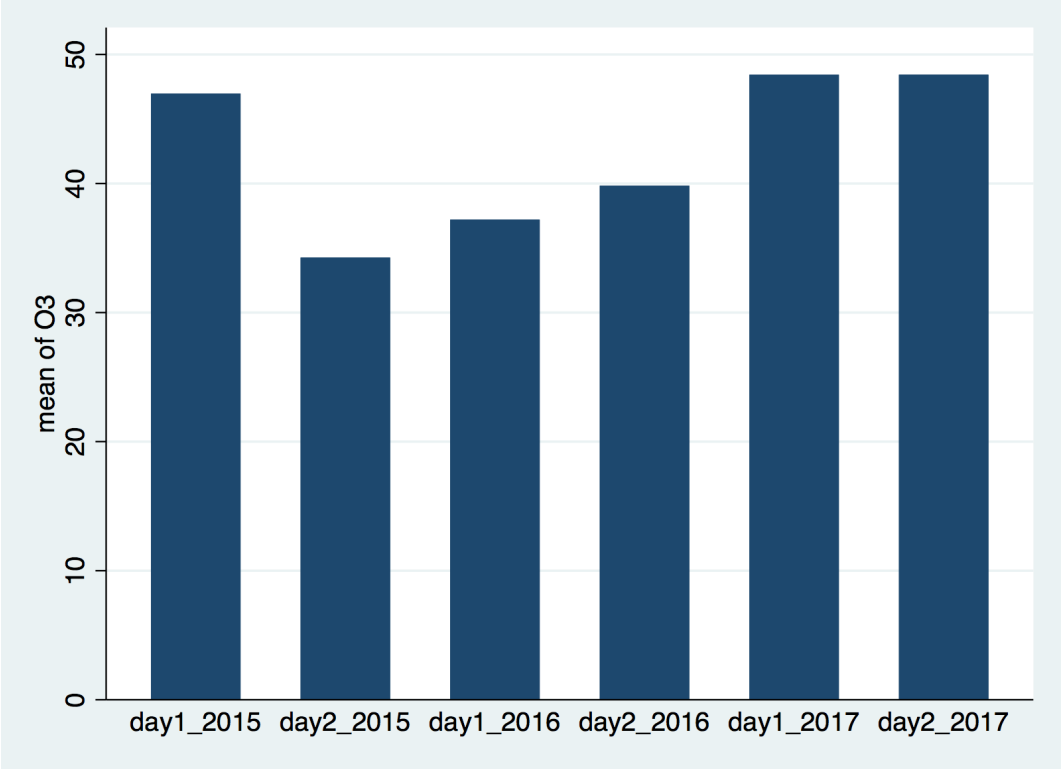
Appendix

Table B1: Instrument for Pollution: Wind Direction – First Stage Results

Cluster_direction	(1) PM_{10}	(2) $PM_{10} > 20$
clus1_90	0.00	0.41
	(0.63)	(0.24)
clus2_90	1.79***	1.43***
	(0.49)	(0.45)
clus3_90	-1.04**	0.12
	(0.44)	(0.3)
clus1_180	0.08	0.22
	(0.32)	(0.27)
clus2_180	1.3***	1.13***
	(0.38)	(0.36)
clus3_180	-0.26	0.62**
	(0.45)	(0.26)
clus1_270	-0.34	-0.02
	(0.23)	(0.2)
clus2_270	1.25**	0.97**
	(0.54)	(0.49)
clus3_270	-0.44	0.35
	(0.44)	(0.25)
Students FE	Y	Y
Day FE	Y	Y
F (First Stage)	55.4	71
Observations	2,405,938	2,405,938

Note: First stage estimates from main specification, using wind direction as an instrument to predict PM_{10} ($10\mu g/m^3$) in column 1; and for the dummy that takes 1 if $PM_{10} > 20$ in column 2. The instrument is the average wind direction in the day that the wind is blowing from each of 4 quadrants. There are 4 directions and 3 regions, hence we have (4-1)*3=9 instruments. Flexible weather controls. Standard errors clustered by municipality and bootstrapped with 500 repetitions.*** p<0.01, ** p<0.05, * p<0.1

Figure B1: O_3 mean per day - IWD - 10km



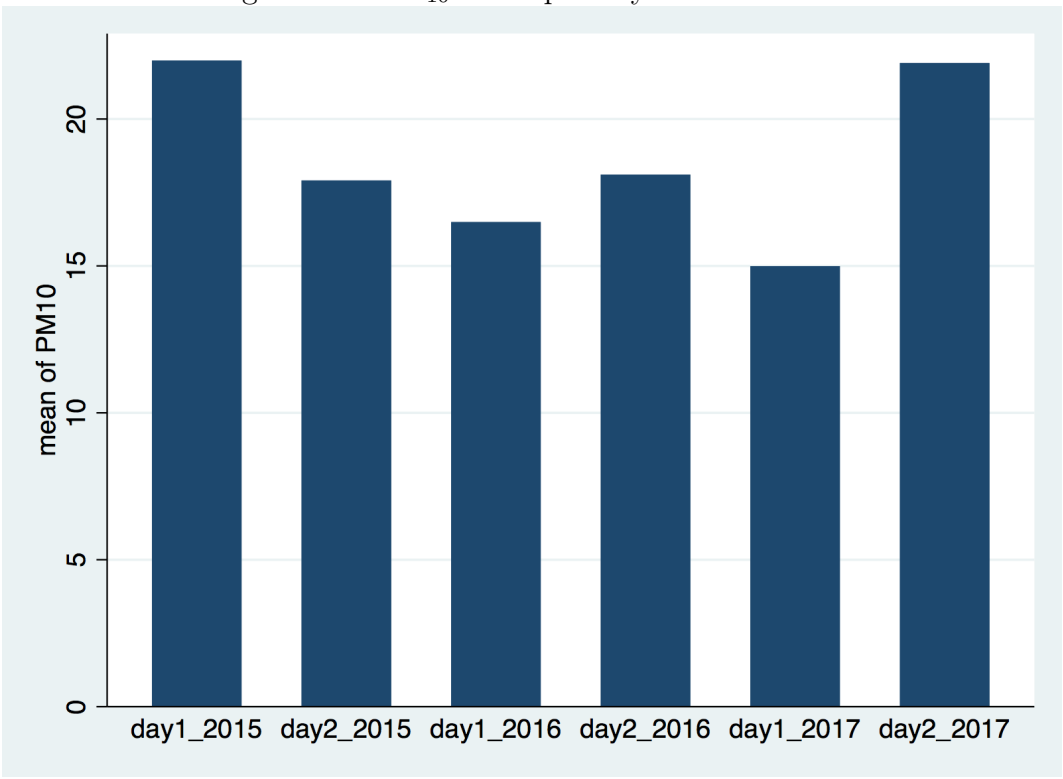
Note: This bar graph plots the average O_3 readings per day of all stations within 10 km from the municipalities centroids using the IWD.

Table B2: The Effects of PM_{10} on Students' Scores (Specification with flexible weather controls)

Variables	(1) Scores (in SD)	(2) Scores (in SD)	(3) Scores (in SD)	(4) Scores (in SD)
PM_{10}	-0.00		-0.03**	
$(10\mu g/m^3)$	(0.04)		(0.01)	
1 if $PM_{10} > 20$		-0.09*		-0.06*
		(0.05)		(0.03)
Students FE	Y	Y	Y	Y
Day FE	N	N	Y	Y
Weather controls	Y	Y	Y	Y
Observations	2,931,368	2,931,368	2,931,368	2,931,368
R-squared	0.04	0.04	0.07	0.07

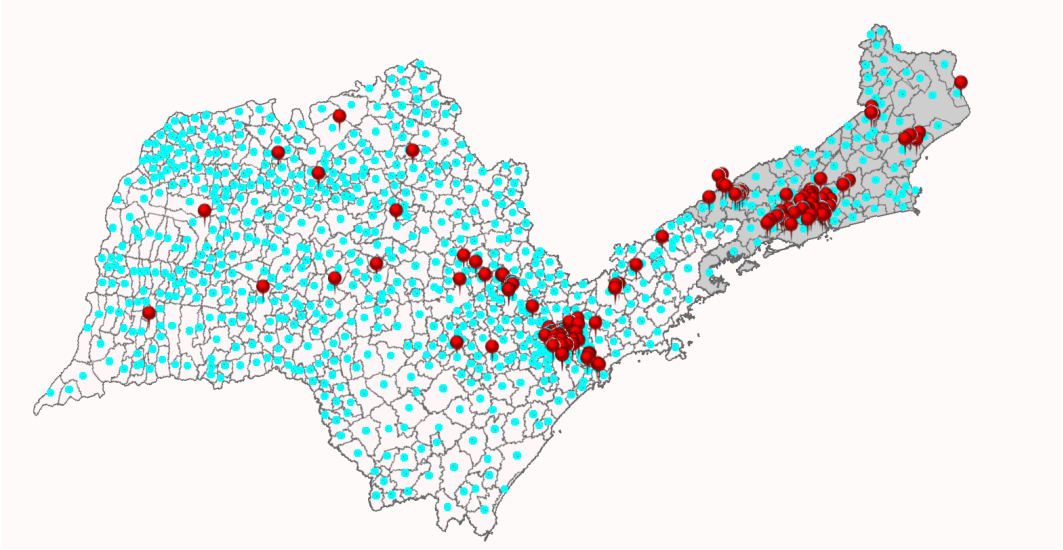
Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE (columns 1 and 2: there is no day FE), temperature and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Figure B2: PM_{10} mean per day - IWD - 10km



Note: This bar graph plots the average PM_{10} readings per day of all stations within 10 km from the municipalities centroids using the IWD.

Figure B3: Rio and São Paulo municipalities centroids and pollution stations (raw data)



Note: The figure plots Rio de Janeiro's (in grey) and São Paulo's (in white) maps at the municipality level. The circles represent the municipalities centroids; and the pins, the location of the pollution stations from our raw data set..

Table B3: The Effects of PM_{10} on Students' Scores (Linear and no weather controls)

Variables	(1) Scores (in SD)	(2) Scores (in SD)	(3) Scores (in SD)	(4) Scores (in SD)
PM_{10}	0.06**		-0.03	
$(10\mu g/m^3)$	(0.03)		(0.03)	
1 if $PM_{10} > 20$		0.08***		-0.04
		(0.03)		(0.04)
Students FE	Y	Y	Y	Y
Day FE	Y	Y	Y	Y
Weather controls	N	N	Y	Y
Observations	3,067,690	3,067,690	2,931,368	2,931,368
R-squared	0.03	0.03	0.06	0.06

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE, linear temperature, relative humidity and wind speed (columns 1 and 2: no weather controls). PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Table B4: Number of municipalities per distance from the monitoring stations

Distance (in km)	Municipalities	Observations
10	36	2,931,368
20	82	4,106,890
30	112	4,560,570
40	132	4,756,535
50	152	4,969,842

Notes: This table displays the number of municipalities and observations in samples that used 10, 20, 30, 40 and 50km distance between the closest monitoring station and the municipality centroid.

Chapter 4

Fetal exposure to fine particulate matter and students' cognitive performance: Evidence from agricultural fires in Brazil

4.1 Introduction

Controlled agricultural burning is a traditional technique still used by farmers, mainly in the developing world, to clean the field from previous harvests and prepare the soil for the next seeding phase. There is growing evidence, however, that the smoke from agricultural fires contributes to increasing levels of fine particulates in the air, causing, as a result, harmful health outcomes for nearby communities (Andreae and Merlet, 2001; Zhang et al., 2016; Chen et al., 2017a). As cognitive performance, via reductions in health, may also be harmed by air pollutants (Almond et al., 2018), investigating the impact of air pollution from these fires on cognitive performance and on future lifetime earnings is quite informative for the outline of agricultural and environmental policies. The purpose of this chapter is to estimate the long-run causal effects of fetal exposure to seasonal air

pollution from agricultural fires on nearby students' academic performance in the national examination in the 5^{th} grade in São Paulo, Brazil.

Despite a growing strand of research showing the short-term negative effects of agricultural fires on cognitive outcomes of the elderly (Lai et al., 2018), or its contemporaneous impacts in reducing students' scores at university entrance examinations (Graff-Zivin et al., 2020), to the best of our knowledge, this is the first study to quantify the long-run causal effects of prenatal exposure to air pollution from seasonal agricultural burning on pupils' scores and wages later in life. To fill this gap, we face a few challenges. Firstly, air pollution levels may be correlated with students' performance owing to their location choice (Bharadwaj et al., 2017) since parents may choose their residence according to the air quality of the region. If this happens, wealthier mothers will have their pregnancy in cleaner areas and, hence, their infants will be exposed to less air pollutants than children born to mothers living in more deprived and usually more polluted localities. Secondly, as seasonal agricultural fires increase both farmers' productivity and revenue, it may affect students' academic performance since the overall household income increases, allowing parents to invest more in their children. Thirdly, sugarcane is harvested in São Paulo in the winter, hence one would expect that some seasonality in weather conditions and winter-related diseases could raise spurious correlation between fires and fetal health. Although a number of studies in the medical literature show a negative effect of biomass burning on health (Arbex et al., 2004; Cançado et al., 2006; Ribeiro, 2008; Uriarte et al., 2009), or its positive impact on asthma hospitalisations (Arbex et al., 2007) in Brazil, these studies still fall short of showing a causal relation by failing to disentangle the confounding effects of other health determinants that are also correlated with controlled fires' smoke.

To tackle the endogeneity problem inherent in linking air pollution to health, and ultimately cognitive performance, we use the seasonal variation of agricultural burning

spots to assess the difference in air pollution. It turns out that, as explained above, fires can also be related to air quality measures. Thus, we follow Rangel and Vogl (2019) and interact fires with wind direction as the latter is able to predict spatial variation in the dispersion of smoke without being directly related to prenatal health or students' scores. As such, we are capable of causally evaluating the different cognitive outcomes of sugarcane burning on pupils living upwind and downwind from the fires. Our study adds to this previous work in three ways. First, Rangel and Vogl (2019) examine the short-term effects of prenatal exposure to agricultural fires on babies' health, while we look at the medium-run (pupils' scores in the 5^{th} grade) and at the long-term (wages). Second, we are interested in the effects of air pollution driven by those fires on human capital. That is, we show the mechanism through which fires affect human health. To this end, we adopt a different econometric approach by utilising, as an instrument, counts of fires upwind from the municipality's population centroid where the student attended school, which we consider being capable of generating plausible exogenous shocks to air pollution [1]. Finally, we use the fine particulate matter ($PM_{2.5}$)[2], which are recognised as having a stronger negative link with health and human capital than PM_{10} used in previous work.

Our study is also related to Graff-Zivin et al. (2020), who investigate the effects of fires' smoke in the day of the Chinese university entrance examinations on candidates' performance, though we offer two more contributions as we pay heed to the medium and long-terms effects of air pollution, while they look only at the contemporaneous impacts. In addition, we complement their study as we are able to establish the causal relation between air pollution and students' cognitive outcomes, an aspect that Graff-Zivin et al. (2020) are not able to examine due to a lack of pollution data for the exact moment of the exams. Also, regarding gender heterogeneity, our work provides evidence that boys' cognitive performance is more impaired by poor air quality than girls', again an aspect that

[1] Unfortunately, we do not have data either on the exact place of birth or on migration. Therefore, we use each municipality's population centroid as a proxy for municipality of birth.

[2] $PM_{2.5}$ refers to fine aerosol particles with a diameter of less than or equal to 2.5 micrometers (μm) near the surface.

Graff-Zivin et al. (2020) lacked data for. More generally, this chapter adds to an extensive body of literature that relates the impact of air pollution to negative health outcomes (Seaton et al., 1995; Chay and Greenstone, 2003; Currie and Neidell, 2005; Ebenstein et al., 2017); to a decrease in productivity of low and high-skilled workers' (Chang et al., 2016; Heyes et al., 2016; Archsmith et al., 2018; Chang et al., 2019; Kahn and Li, 2019); to an increase in school absenteeism (Currie et al., 2009; Ransom and Pope, 2013; Chen et al., 2018; Liu and Salvo, 2018); and to a decrease in cognitive performance (Currie et al., 2009; Zweig et al., 2009; Reyes, 2011; Ebenstein et al., 2016; Bharadwaj et al., 2017) due to its harmful effects on the respiratory and cardiovascular systems (Dockery et al., 1993; Seaton et al., 1995), insofar as we aim to demonstrate the medium and long-run damages caused by air pollution derived from agricultural fires during pregnancy in locations where $PM_{2.5}$ measures are not as high as in the vast majority of other studies. In addition, our pollution source is not only delimited but also within an agricultural setting, differing from other already explored sources, such as vehicles, plants, chimneys, or airports; thus, expelling different chemical substances to the atmosphere. To put it into perspective, we calculate that around 61% of fires take place in agricultural lands within our data set[3]. As such, we shed light on the important fact that, even at relatively moderate levels, this pollutant has the potential to cause cognitive ability impairment.

Some studies have shown that exposure to Ozone, and to the pollutants it stems from (NOx), is related to premature death, asthma, bronchitis, heart attack, and other cardiopulmonary illnesses (Weinhold, 2008; Currie et al., 2009; Neidell, 2009; Developers, 2016; Wikipedia, 2019), and, even at low levels, more susceptible individuals are more likely to be affected by this pollutant (Bell et al., 2006; Weinhold, 2008). Regarding SO_2, this pollutant has a number of harmful health impacts, including breathing difficulty, pulmonary edema, eye irritation, asthma attacks, cardiopulmonary diseases and increase of

[3]The rest is spread across urban areas (3.2%), grazing(15%), forestry(15.3%), with the remaining fires being spotted in forests and other fields. To make these calculations, we overlay our fires data from NASA/FIRMS with land coverage data from the agricultural census from 2010 available at the IBGE website.

mortality rates (Lin et al., 2004; Khan and Siddiqui, 2014; Goudarzi et al., 2016; Khaniabadi et al., 2017). As the brain requires large amounts of oxygen to function well, medical studies have suggested that the cognitive performance is likely damaged as a result of the exposure both in the short and long-run to air pollutants (Calderón-Garcidueñas et al., 2008). The medical literature shows that different parts of the brain perform different functions, with white matter being linked to language and verbal reasoning skills and gray matter associated with mathematical skills (Popescu et al., 2019; Houston et al., 2019). Other works associate exposure to air pollution with a decrease in total gray and white matter volumes (Erickson et al., 2020). Based on this evidence, the economics literature has suggested a causal link between pollutants and a deficit in cognitive function (Currie et al., 2009; Zweig et al., 2009; Reyes, 2011; Ebenstein et al., 2016; Bharadwaj et al., 2017).

We focus our attention on sugarcane since the bulk of the fires take place in areas dedicated to this crop. Nonetheless, there are others crops that also use the burning technique, such as maize, rice and wheat. One can observe from Figure C8 the share of sugarcane production per total cultivated land in all municipalities in São Paulo and, comparing to Figure C9, sugarcane plantation has increased since 2000, shifting the other three products away. Furthermore, when it comes to relating the occurrence of fires with sugarcane, we can notice from Figures C1 and C3 that the areas with more fires detected coincide with the hot spot areas of this crop. Sugarcane fires take place in large areas emitting aerosols, such as poly-cyclic aromatic hydrocarbons (PAHs), organic and black carbon, sulfur dioxide (SO_2), ozone (O_3), nitrogen oxides (NOx) and particulate matters ($PM_{2.5}$ and PM_{10}) and is carried by wind towards surrounding localities, which usually are very populated (Guoliang et al., 2008; Akagi et al., 2011; Wu et al., 2018). This feature makes straw burning a rather suitable source of air pollution given that not only do we know the exact location of the fire points, but also the fact that agricultural fires are close to populated regions, raising concerns about human capital damages much more than if we were exploring wild fires, which generally take place in remote lands. Last

but not least, this chapter complements Bharadwaj et al. (2017) and Isen et al. (2014) in calculating the returns of cleaner air quality on scores and, consequently, on wages (long-run effects) in the largest Latin American country.

With regards to our empirical design, we build an individual-level pooled cross-sectional data set, collected from the Ministry of Education of Brazil, with scores from students born between 2001 and 2008, which we merge by month, year of birth and municipality of residence with the exact location of the active fires. The latter data come from the National Aeronautics and Space Administration (NASA)'s active fires product. Besides that, we use daily atmospheric conditions from satellite remote sensing re-analysis gridded data by the Japanese 55-year Re-analysis (JRA-55), from which we utilise the average wind direction per day with or without fire to create our instruments. These consist of the counts of fires per wind direction: upwind, downwind and vertical wind. That is, we consider upwind fires those whose smoke would blow towards a certain municipality's population centroid; downwind fires are those in the opposite direction; and vertical are fires in all the remaining directions. Our first stage results show that, under our instrumental variables approach, our main pollutant measure, $PM_{2.5}$, can be considered randomly assigned since we find stronger and more significant correlation between upwind fires and pollution than other fires. In addition, our instruments of fire counts are built in several ways, and proven to be strong for $PM_{2.5}$. To link students' scores to air pollution level *in utero*, we use satellite gridded data from NASA's second Modern-Era Retrospective analysis for Research and Applications (MERRA-2) product, merging it by municipality-month-year of birth.

Our results indicate the external costs borne by individuals stemming from agricultural burning inasmuch as it impedes the capital formation and, ultimately, the economic growth of the country (Solow, 1999). This is of particular importance when it comes to developing countries, where the returns to education are generally higher than developed

nations, making São Paulo an interesting case study since the state yields around one fifth of the global production of sugarcane, with the largest production within the country and in the world (FAOSTAT, 2020; Unicadata, 2020). In response to a national and international rise in demand for bio-fuel and sugar, the state more than doubled from 10% to 23% (Figures C1 and C3 offer a visualisation of this) its share of area designated to sugarcane production throughout the past two decades, containing two thirds of all Brazilian production (McConnell et al., 2010; Rangel and Vogl, 2019). Moreover, despite the relative decline in the number of agricultural fires in the past years in Brazil, due to become illegal from 2021 with the Green Ethanol Protocol in 2017 signed by the government (SIMA, 2017), controlled burning is still a technique substantially applied in the country, with burning crops starting to see its peak in our period of study as Figures C5 and C7 illustrate. The harvesting season for sugarcane starts in April until November, when the farmers who use the manual method - instead of the mechanical - set controlled fire to the fields to crop the cane, increasing the labour productivity had the method not been applied (Fernandes, 1988; Kirchhoff et al., 1991; Rangel and Vogl, 2019). Usually the fires take from a few hours to 24 hours during the peak (Lamsal et al., 2017).

In this setting and consistent with existing research, our findings suggest that an increase of 10μ/m3 of $PM_{2.5}$ during pregnancy decreases students' Portuguese scores by 4.6% and Maths scores by 2.8%, with boys more affected than girls. Back-of-the-envelope calculations indicate that a reduction of $PM_{2.5}$ in the third trimester by 2.4 SD, i.e., 10μ/m3, would increase child's scores by 4.6% and ultimately her wages by 2.3%. Our results are robust to a set of sensitivity checks. Therefore, our study provides policymakers with important evidence of a link between prenatal exposure to air pollution and lower cognitive performance.

The remainder of this chapter is organized as follows. In Section 2, we describe the data. In Section 3, we detail the empirical strategy. In Section 4, we present results. In

Section 5, we show sensitivity analysis, and Section 6 concludes.

4.2 Data

We build a repeated individual-level cross-sectional data set on monthly air quality, weather conditions, agricultural fires and students' scores in the fifth grade in São Paulo. We choose to examine cohorts of students born between 2001 and 2008 because some of our remote sensing data set come from the satellites Terra and Aqua, which started to operate between 1999 and 2002. Furthermore, we opt for remote sensing data instead of ground-based monitoring stations since the agricultural fires are eminently located within rural areas, in which very scarce pollution and meteorological stations could be found for our time period of analysis. Our data are collected from four main sources as detailed below.

4.2.1 Agricultural Fires

We collect daily remote-sensing data on agricultural fires from the National Aeronautics and Space Administration (NASA)'s Fire Information for Resource Management (FIRMS). The data are captured by two satellites TERRA and AQUA, which rely on Moderate Resolution Imaging Spectroradiometer (MODIS) sensors to detect ground-level fire activity. Each satellite overpasses Brazil twice per day: in the morning, afternoon, evening, and late nights (Kaufman et al., 1998; Justice et al., 2002). The fire detection algorithm developed by researchers at the University of Maryland is built upon thermal anomalies, surface reflectiveness, water, cloud masking and land use, and the publicly available data is provided with a resolution of 1km X 1km (Giglio et al., 2009, 2016). Notwithstanding, it is not possible to know the precise size of each fire from the data.

The product also gives the confidence of detection for each potential fire identified by the satellites. That is, each fire receives a probability of real occurrence (from 0% to 100%) according to the meteorological conditions and vegetation at the exact time the fire was detected (Giglio et al., 2016). As a result, in our specifications we use the fire weighted by its confidence to minimise classical measurement error (Rangel and Vogl, 2019; Graff-Zivin et al., 2020).

A fire is attributed to a student if it occurred within a catchment area of 50km radius of her municipality's population centroid. Unfortunately, we do not know the municipality of birth of each individual. Hence, we use the municipality's population centroid where the pupil lives when she takes the exams as a proxy for the exact location of birth. Nevertheless, alternative distances are used and presented in the robustness checks section.

To build our fire variables, we follow Rangel and Vogl (2019) and count the daily number of grid cells containing a fire within a 50km of a municipality's population centroid. Then, we consider that fires detected within the same wind octant in a day are upwind, fires located in the opposite octant are downwind, and all the remainder fires are defined as vertical. As in previous works (Rangel and Vogl, 2019; Graff-Zivin et al., 2020), for our main specification, we set a 45^0 angle. As such, we define upwind fires as those located within a 45 central angle from the average daily wind direction within each municipality. In other words, upwind fires have the smoke blowing towards the municipality's population centroid. Downwind fires are located in the opposite direction; and we call vertical fires the ones situated in the other directions. The rationale of this classification is that municipalities situated upwind from fires are exposed to more air pollution coming from the fires than downwind locations, whereas the air pollution in the municipalities located in the vertical direction is less influenced by smoke than the upwind villages, but more affected by fires than their downwind counterparts. In our sensitivity analysis section we aggregate vertical and downwind directions to form an alternative set of instruments

since these directions present similar behaviour. Moreover, we run models using upwind and downwind fires, excluding the vertical ones, and the results remain very close to the main specification.

In Table 4.1, we report the summary statistics for agricultural fires per wind direction used for the whole period of the main estimation sample (2001-2008). Also, Figures C5 and C6 present the time series behaviour of monthly average fires for that period. From these figures we can observe a strong seasonal pattern of fires with peaks in the winter months, and a not very clear time trend. The seasonality of fires coincides with the agricultural phases, corroborating the link between sugarcane harvesting period and the peak in identified burning spots. In Figure C7, we present the overall time series behaviour from 2001 until 2017; i.e., we add the period ranging from 2009 to 2017 since the students took exams from 2011 to 2017.

4.2.2 Atmospheric Conditions

Re-analysis Data

We use the Japanese 55-year Re-analysis (JRA-55) data on atmospheric conditions. The JRA-55 is the second Japanese global atmospheric re-analysis project conducted by the Japan Meteorological Agency (JMA), covering data back to 1958.

Re-analysis uses past climatic observations from satellite and terrestrial monitoring stations, as well as meteorological models to systematically produce data sets for climate monitoring and research. On one hand, the main advantages of working with re-analysis data are: i) their consistent spatial and temporal resolution and baseline models which have reduced biases over each product's new edition; and ii) re-analysis inputs millions of observations that would be impossible for a researcher to collect and deal with it separately. On the other hand, the key limitation is the biases in observations and models that

lead to unavoidable classical measurement error. Hence, the re-analysis data and results of models using them should be interpreted with caution.[4] Nevertheless, to attest the scientific validity of this data set, several works in the Atmospheric Sciences field evaluated the precision of the JRA-55 products in comparison to others using different models by different researchers, finding that these re-analysis products lack of perfect accuracy but are still useful for climate studies (see for example, Sun et al., 2018; Tsujino et al., 2018).

Weather Conditions

We gather gridded data on temperature in Kelvins, which we transform to degrees Celsius (^0C), precipitation (in millimeters, mm), relative humidity (in percentage, %), wind components and cloud coverage (in percentage, %). Apart from the latter, these weather conditions may influence the way air pollution geographically spreads. As such, we include them as control variables in our models. Regarding cloud coverage, it is important since the satellite imagery for fire detection is negatively related to the amount of cloud in the sky.

The spatial resolution of each grid is 0.56° X 0.56°, and the temporal is 3-hourly data that we average by day per grid. We then calculate the monthly average at the grid level per occurrence of fire. That is, we find the monthly mean by aggregating the days with fires and the days without registered fires since weather conditions may different in days with and without fires. By doing so, we aim to mitigate classical measurement error had we imputed weather conditions from days without fires for days with fires and vice-versa.

To construct the variables for wind conditions we use the East-West wind direction (u-component) vector and the North-South (v-component) vector. The wind speed (in

[4]For more details on how re-analysis data are produced, please, visit the website https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables.

meters per second, m/s) is calculated by:

$$Wind\ speed = \sqrt{u^2 + v^2} \tag{4.1}$$

To calculate the average wind direction for each observation, firstly, we transform the degrees into radians; secondly, we compute the North/South direction = y = cosine(radians) X wind speed; the East/West direction = x = sine(radians) X wind speed; sum all ys (Y) and sum all xs (X); calculate the arctangent (the inverse tangent) using Y and X. The results are finally transformed back from radius to angle (Webmet, 2019; Itron, 2019). One should note that the average wind direction refers to the direction the wind is blowing from since we are interested in the direction of the wind in relation to the fire point blowing towards the municipality's population centroid.

4.2.3 Air Pollution

The pollutants and other chemical elements, such as $PM_{2.5}$, SO_2, Sulfate (SO_4), O_3, Black Carbon (BC) and Organic Carbon (OC), dust (DS) and sea salt (SS), utilised in this chapter come from the second Modern-Era Retrospective analysis for Research and Applications (MERRA-2), provided by NASA and built upon a plethora of data sources, such as remote sensing imagery and terrestrial stations. The gridded data available has a spatial resolution of 0.5° X 0.625° and comes in hourly intervals, which we aggregate to daily and monthly frequency. To this end we implement the same procedure as for the weather variables; i.e., we first average per day distinguishing days with or without fires, then we calculate the monthly average.

$PM_{2.5}$ is not directly available from the Merra-2's data set. Thus we follow the Atmospheric Science literature to calculate the particulate matter using the formula below

(Buchard et al., 2016; Provençal et al., 2017).

$$PM_{2.5} = 1.375 * SO_4 + 1.8 * OC + BC + DS + SS \tag{4.2}$$

where SO_4, OC, BC, DS and SS are measured in kilogram per cubic meter (kg/m^3), which we finally transform to micro-gram per cubic meter (μ/m^3).

As per any re-analysis data, one needs to be careful when interpreting our results using these pollutants data since their use feasibly could induce some attenuation bias in the econometric estimations. More precisely, measuring pollution via remote sensing still present a few weaknesses which have been assessed by the Atmospheric scientists (Liu et al., 2009; Paciorek and Liu, 2009; Estes Jr et al., 2010). For example, Buchard et al. (2016) and Provençal et al. (2017) assess the precision of satellite-based aerosol optical depth (AODS) in comparison to air pollution monitored on the ground in the United States and in Europe. Both works demonstrate that the AOD figures are quite reliable in relation to the terrestrial stations' readings.

4.2.4 Education

Our data on students' scores come from the National Institute for Educational Research *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP), an agency from the Brazilian Ministry of Education. Every two years the students in the fifth grade pass a nationwide examination to assess the quality of education in Brazil (*Prova Brasil)* via tests and socioeconomic questionnaires (INEP, 2017a). This forms a rich administrative data set with detailed information on the students' scores in Portuguese and Mathematics, demographic information, such as parental level of education, age, race, gender and special needs. From this data we build a repeated pooled cross-section at the student level with cohorts taking examinations from 2011 to 2017 (Table

4.1 summarises this information). The Portuguese scores range from 0 to 350 points, and Maths from 0 to 400 points, and, in our regressions, we follow the literature and use these scores in standard deviation form as dependent variables (Garg et al., 2020). In addition, we include cohorts of students taking the fifth grade exams in 2011, 2013, 2015 and 2017. As for the year of birth, the children in our data set were born from 2001 to 2008.

The INEP data provide the highest level of education achieved by each student's parents. As such, we first drop the missing observations in which the students declared they were unaware of their parents' level of education. Then, we create six dummies for each parent's education: (i) never studied, (ii) incomplete 1^{st} elementary if the parent attended school for a while, but did not complete the first elementary, (iii) incomplete 2^{nd} elementary if the parent attended school for a while, but did not complete the second elementary, (iv) elementary if the parent only completed elementary School, (v) high school if the parent completed the high school, and (vi) undergraduate college education (Zweig et al., 2009). To avoid the dummy trap, we exclude one dummy for each parent's education. We also control for students' race and gender.

In Table 4.1, we present the descriptive statistics for our main estimation sample with 1,190,717 students. The average daily concentration of $PM_{2.5}$ in our estimation sample is 14.3 $\mu g/m^3$, with a standard deviation of 4.24. For SO_2 the respective figures are 4.06 $\mu g/m^3$ and 2.82.

Table 4.1: Summary Statistics for main regressions sample

Variables	Mean	SD	Min	Max	Obs
Scores					
Portuguese	219	47.8	84.9	339	1,190,717
Maths	234	47.5	78.92	367	1,190,717
Pollutants					
$PM_{2.5}$ ($\mu g/m^3$)	14.3	4.24	4.64	59.5	1,190,717
SO_2 ($\mu g/m^3$)	4.06	2.82	0.11	28.8	1,190,717
Agricultural Fires					
Upwind	8.61	12.8	0	129	1,190,717
Downwind	7.49	10.5	0	125	1,190,717
Vertical	14.5	20.8	0	191	1,190,717
Meteorological Conditions					
Rain (mm)	2.73	2.66	0.00	92.4	1,190,717
Temperature (^0C)	22.2	2.69	13.4	35.6	1,190,717
Wind Speed (m/s)	2.05	0.77	0.13	8.52	1,190,717
Cloud coverage (%)	40.4	13.1	6.24	99.9	1,190,717
Humidity (%)	70.3	8.54	25	97.6	1,190,717

Notes: Summary statistics of some of the key variables: scores ranging from 0 to 350 points for Portuguese and 0 to 400 points for Maths, counts of agricultural fires and meteorological conditions during pregnancy (from 2000 to 2008) in the municipalities in the State of São Paulo. The fires locations are defined within 45 degrees from the daily average wind direction in a municipality.

4.3 Empirical Strategy

4.3.1 The basic OLS models

We examine the partial correlation between $PM_{2.5}$ fetal exposure and test scores in the 5^{th} grade by pooling our individual-level cross-sectional data for the exams happening in years 2011 to 2017. The basic specification takes the following form:

$$Y_{jmit} = \phi pollutant_{mi} + \psi W_{mi} + \alpha X_j + \mu_m + \iota_i + \theta_t + \epsilon_{jmit} \tag{4.3}$$

where Y_{jmit} is the 5^{th} grade test score in either Portuguese or Maths of child j born in municipality m, at time i, taking exam at year t. The parameter of interest is ϕ, i.e., the coefficient of $PM_{2.5}$ measuring the average level of pollution at the trimester and

municipality levels. Our study focuses on the effect of $PM_{2.5}$ on cognitive performance, although in robustness exercises we include SO_2 as a co-pollutant. W represents a vector with weather conditions controls during pregnancy, such as temperature, relative humidity, wind speed, cloud coverage, and rainfall per trimester. Since temperature extremes can affect fetal health and air pollution concentration, suggesting non-linearity in its effects, we create five temperature bins between less than 15 and more than 35 degrees (5 degree bins) in our main specification, but also, for completeness, experiment with using temperature in the linear form 4.8, squared, and interacted with other weather variables (Deschênes et al., 2009; Bharadwaj et al., 2017). X_j is a vector with controls for observable individual characteristics that can be related to their exam scores, such as the mother's highest level of education[5], students' race and gender. μ_m are municipality fixed effects, we include ι_i, a vector to control for month and year of birth fixed effects. The two latter account for unobserved time-invariant municipality and cohort characteristics, and θ_t contains dummies to control for year of exam. The standard errors are clustered by municipality to control for spatial and serial correlation within each locality. Unfortunately, we do not have the exact place of birth, thus we use the municipality where the student attended school as a proxy for municipality of birth.

Some concerns arise in using OLS. The first relates to the fact that students' exam scores may be correlated with air pollution since the wealthier parents could choose to live in less polluted areas (Banzhaf and Walsh, 2008; Currie et al., 2014). This location choice is a form of avoidance behaviour, which may bias our estimations and needs to be tackled. (Neidell, 2004, 2009; Bharadwaj et al., 2017). Second, as seasonal agricultural fires increase both farmers productivity and revenue, it may affect students' academic performance since the overall household income increases, allowing parents to invest more in their children. Third, sugarcane is harvested in São Paulo in winter, hence one would

[5]We do not have the years of education of the student's parents. Rather, the original micro-data provides the highest level of education achieved by them. As such, we build dummies indicating each of these levels. For more details on the construction of the dummies, please refer to section 3.

expect that some seasonality in weather conditions and winter-related diseases could cause a spurious correlation between fires and fetal health. To try to mitigate these challenges we opt by using an instrumental variable approach to estimate our models. The novelty of our work is that we combine agricultural fires with wind direction to predict air pollution.

In addition, our type of pollution measure and meteorological data set, which are produced by re-analysis modelling, may bias our estimates inasmuch as children's actual exposure to outdoor air pollution may differ from the forecast readings in our data. Hence, our instrumental variable strategy can also mitigate this classical measurement error problem (Moretti and Neidell, 2011; Schlenker and Walker, 2015; Chen et al., 2018). Finally, one could also argue that the level of pollution and weather conditions in the days the exams are marked could influence the scores as the examiners could also have their productivity affected by these circumstances. However, all the tests are multiple choice and the grades are given by a machine, and thus should not be affected by air pollution or weather determinants.

4.3.2 The instrumental variable approach

After using fixed effects and controlling for weather conditions, to identify the causal relation between pollution and students' scores we partially build on Rangel and Vogl (2019)'s strategy of interacting agricultural fires with wind direction to construct our instrumental variables. Our main assumption relies upon the exogeneity of the upwind/downwind fires since it does not directly affect the children's cognitive performance, except from its effect on air pollution. As such, our instrument can be considered as a natural experiment that results in exogenous shocks to local air pollution in a specific area (Schlenker and Walker, 2015; Anderson, 2015; Deryugina et al., 2019b; Graff-Zivin et al., 2020).

The interaction of fires with wind direction is due to the concern that agricultural fires may be correlated with students' academic performance insofar as areas that use this technique can present higher returns in productivity in comparison to other locations which do not use fires to accelerate the harvesting process. Thus, in light of the possibility of agricultural fires not fulfilling the exogeneity requirement for a good instrument, we interact this variable with wind direction, which influences the spread of the fire's smokes - and as a result air pollution - but does not affect students' scores.

To deal with the endogeneity concerns above, we take advantage of the influence of the temporal and spatial variation in wind direction on the spread of the fires' smoke, and hence on air pollution levels in the nearby municipalities to construct our instruments. To this end, we count the amount of fires upwind and downwind up to 50km from the municipalities' population centroids during each of the three trimesters of pregnancy. Following Rangel and Vogl (2019), we define the wind directions using the 45 degrees.[6] In our main specification, we include the counts of upwind and downwind fires per trimester divided by the amount of days of each period. In two robustness exercises we also include the three directions, i.e., upwind, downwind and vertical wind, or sum up the counts of fires downwind with vertical as a unique category along with upwind fires.

The underpinning assumption of our instrumental variable approach is that fires located upwind from the child's municipality of birth elevate pollution more than downwind fires. As such, in this section, we present the reduced form equation to estimate the effects of agricultural fires on air pollution and the final impact of the latter on scores. The count of upwind fires is a strong predictor of local $PM_{2.5}$ levels as Table 4.2 shows. Therefore,

[6] Graff-Zivin et al. (2020) explains that there is a trade off between capturing more fires per angle (using wider angles) and more precision in determining wind direction (using narrower angles). The authors show that 45 degrees is the octant that best captures pollution dispersion and are followed by Graff-Zivin et al. (2020).

our model can be formally described as:

$$Pollutant_{mi} = \gamma upwindfires_{mi} + \zeta downwindfires_{mi} + \psi W_{mi} + \alpha X_j + \mu_m + \iota_i + \theta_t + \epsilon_{mni} \tag{4.4}$$

$$Y_{jmit} = \beta \widehat{Pollutant}_{mi} + \omega W_{mi} + \rho X_j + \nu_m + \lambda_i + \delta_t + \eta_{jmit} \tag{4.5}$$

where, in both equations, the variables are the same as in Equation 4.3, except for the outcome variable in Equation 4.4, which here is the pollutant measure linked to students j, in municipality m and time i, taking exams in t (Eq. 4.4), and the main explanatory variable in equation 4.5, which is the predicted value from the reduced form (Eq. 4.4). The coefficient of interest in equation 4.4 is γ, which informs the estimated effect of upwind fires on $PM_{2.5}$. The pollution level and weather controls correspond to the monthly average during each of the three trimesters of pregnancy, and the fire counts, to the amount of fires in each trimester. In Equation 4.5, we are interested in β as it reports the estimated causal effects of pollution on pupils' scores in the years of the examination (2011 to 2017). The other control variables, such as students' characteristics and parents' level of education, refer also to the time of the tests. We add municipalities fixed effects to control for unobserved factors that do not vary over time, for year of examinations and year and month of birth as explained before.

The literature discusses the possibility of avoidance behaviour by parents during pregnancy. For example, parents may change their behaviour in locations where there exist pollution levels alerts (Neidell, 2006, 2009; Graff-Zivin et al., 2011; Moretti and Neidell, 2011; Bharadwaj et al., 2017). Despite that, we are not aware of air pollution alerts system within our study region. As for other types of avoidance behaviour, such as closing windows, it is not possible to account for with our data.

Lastly, a limitation for studies on fetal shocks effects later in life is how to address the influence of time-varying investments parents may make in human capital to overcome fetal exposure to pollution until the exams years, which is not possible to control for in our sample (Bharadwaj et al., 2017; Almond et al., 2018). Nonetheless, other works have shown that the compensating parental responses present very small effects (Bharadwaj et al., 2013; Halla and Zweimüller, 2014; Bharadwaj et al., 2018). Therefore, if there is still some sort of unobservable investments that parents make to compensate cognitive impairments stemming from in utero pollution exposure, they will be contained in our error term and represent another negative externality from air pollution revealed as additional parental costs.[7]

4.4 Results

4.4.1 Mechanisms: Relevance of upwind fires as instrument

From Table 4.2 one can see that upwind fires are strongly causally related to the level of air pollution per trimester of pregnancy. Columns 1 to 3 refer to equations including upwind and downwind fires, while Columns 4 to 6 show results for specifications using downwind and vertical wind in a same category, and Columns 7 to 9 contain the results from the specifications which include the three categories of wind. Since we have three endogenous variables - the average $PM_{2.5}$ by trimester - we need three first stage equations. Moreover, we interact each wind direction with fires also per trimester. Therefore, there are enough instruments to support three endogenous variables within the same equation.

From the results one can see that upwind fires persistently affect pollution more than the other directions across the different forms for which we build the instruments set. This reinforces the main assumption that we rely on that local air quality is more affected

[7] Almond et al. (2018) reports a thorough review of the literature on fetal shocks.

by fires located upwind from the municipality's population centroid. In addition, and in line with Lai et al. (2018) and Graff-Zivin et al. (2020), we confirm that poor air quality is the channel through which agricultural fires affect humans' cognitive health.

4.4.2 Second Stage: Impact of air pollution on students' scores

Table 4.3 reports the results regarding the effects of $PM_{2.5}$ per three trimesters of pregnancy on students' scores. All models include flexible weather controls, and the equations estimated by Two Stage Least Squares (2SLS) include the share of fires (weighted by their probability of occurrence) per wind direction. We start by presenting our OLS estimates with Portuguese (Panel A Columns 1 and 2) and Maths (Panel B Columns 1 and 2) scores. These results show that $PM_{2.5}$ in the first and second trimesters seem to have a positive impact on Portuguese and on Maths, and no effects on the third gestational period. However, for the reasons explained above, OLS estimates are likely biased. For example, if the students are from poorer municipalities and more polluted areas, OLS may overestimate the effects of air pollution on their scores. On the other hand, if the student comes from a wealthier and polluted area, parents' compensating behaviour may not be captured by that estimator, resulting in a underestimated pollutant's coefficient Ebenstein et al. (2016). We present results using instrumental variables with and without controlling for students' characteristics (Columns 3 and 4). Comparing both columns, one notice that the magnitude of the estimates does not seem to change much by including students' controls.

We then turn to our main specifications using the counts of upwind and downwind fires as instrumental variables. One can observe that an increase of 10 $\mu g/m^3$ of $PM_{2.5}$ during the third trimester of pregnancy causes literacy reasoning to decrease by 0.21SD (Column 4 Panel A) and numeracy performance by 0.14SD (Column 4 Panel B). With regard to this, the literature shows evidence that different parts of the brain perform different functions with white matter being linked to language and verbal reasoning skills

and gray matter associated with mathematical skills (Popescu et al., 2019; Houston et al., 2019). In addition, pollution affects the white and gray matter differently, corroborating our findings suggesting that air pollution generates slightly different effects on exam performance by subject (Clougherty, 2010; Calderón-Garcidueñas et al., 2011; Erickson et al., 2020).

When we look at the first trimester point estimates, we note that the instrumental variable specification correctly dealt with the signs; i.e., contrary to the OLS likely biased positive estimates, the quasi-experimental identification now shows a negative impact although not significant. In addition, the exposure to this pollutant during the second trimester continues to spuriously suggest that pollution leads to an increase in both tests scores. These two results relating to earlier gestational periods are not as surprising as one might think since the literature points to the existence of selection bias when it comes to backwards dating of exposure to air pollution and the exact date of birth. Thus, this is a recurrent issue in several works on in utero shocks. For example, Rangel and Vogl (2019)'s findings also suffer from the same limitation.

4.4.3 Heterogeneity

In this section we explore the heterogeneous effect of the treatment effects (air pollution) for our preferred specification presented in Table 4.3 (Column 4): (i) per gender and (ii) across the distribution of scores per municipality. Through this heterogeneity analysis our goal is to unveil how different sub-populations have their cognitive performance later in life affected by the fetal exposure to poor air quality.

With regards to gender differentials, overall our findings reported in Table 4.4 suggest that boys are more negatively affected by air pollution than girls. Column 1 repeats the results of our baseline model (the same from Table 4.3) for ease of comparison to girls' and

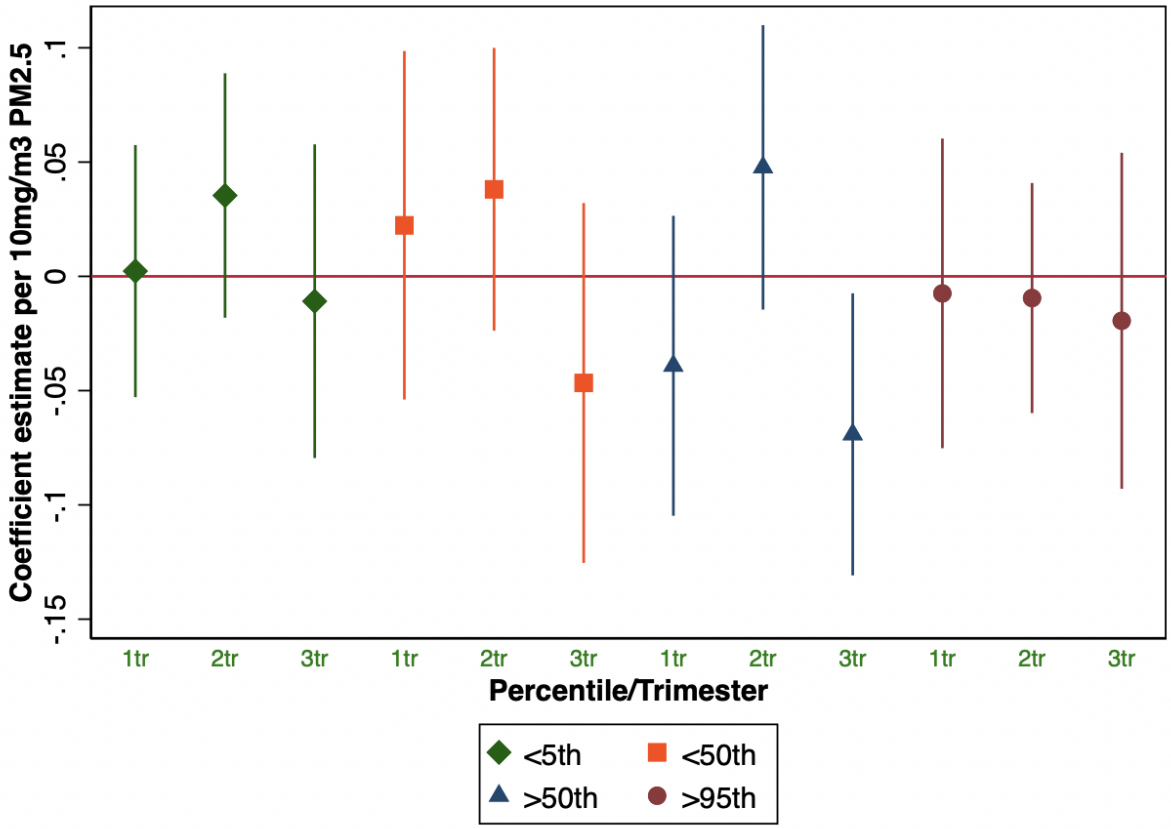
boys' performances (Columns 2 and 3) on Portuguese exams. Firstly, one can note that the pattern of signs and significance concerning the first two gestational periods remain the same, with the exposure to $PM_{2.5}$ during the first trimester showing not significant effects on cognitive performance at school, and the second trimester appearing to be positive. To put this into perspective, when it comes to the third trimester, girls seem to be less impacted by the pollutant than boys, with the latter suffering a decrease of 0.29SD in their Portuguese marks for an additional 10 $\mu g/m^3$ of $PM_{2.5}$, whereas their female counterparts have their scores dropped by 0.12SD with less significance (10%). Turning to Maths scores (Columns 4 (main specification) to 6), the results are similar, with boys' marks decreasing by 0.19SD per a rise of 10 $\mu g/m^3$ of $PM_{2.5}$, and girls' numeracy reasoning not significantly affected by pollution. This finding is consistent with previous studies including Ebenstein et al. (2016) and the results in Chapter 3, which suggest that males suffer a larger drop in their tests' scores when exposed to poor air quality during the examinations.

Although the role of gender in the effects of air pollution on respiratory health remains unclear, some epidemiology studies point out that males are more affected than females. For example, Abbey et al. (1998) link PM_{10} exposure to reduced lung function among males but not females, Galizia and Kinney (1999) find similar results regarding the exposure to O_3, and Chen et al. (2017b) identify that contemporaneous and cumulative exposure to air pollution has a stronger negative effect for men than for women. Similarly, Ebenstein et al. (2016) find the effect of pollution on exam performance to differ by gender and cite the increased incidence of asthma amongst boys as a possible cause.

As per the second heterogeneity analysis, we aim to detect whether strong or middling students are differently affected by poor air quality. To this end we split our final sample into subsets by low and high scores within each municipality with results for Portuguese and Maths scores as dependent variables depicted in Figures 4.1 and 4.2. One

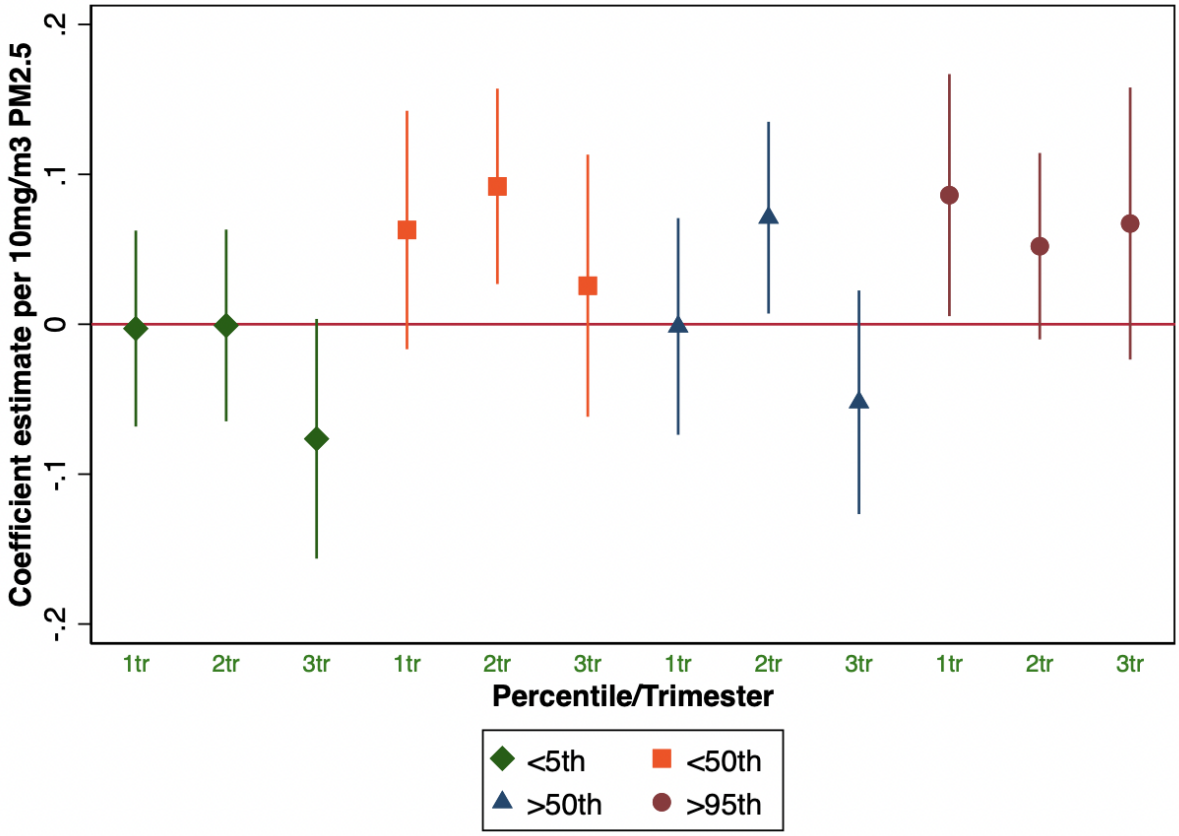
can note that weaker students (5^{th} percentile) are more sensitive to exposure during the third trimester of pregnancy than the others in the scores distributions for both subjects. Nonetheless, students above the median also have their marks diminished as a result of fetal exposure to air pollution during the third period of pregnancy. Noteworthy is the fact that top students within the 95^{th} percentile seem also to have their Maths' scores reduced. Both results are in line with Ebenstein et al. (2016) who also show evidence of weaker students' poorer performance as a result of bad air quality during the university examinations, and Graff-Zivin et al. (2020) who show that students belonging to the top of the distribution are more affected by bad air quality.

Figure 4.1: Heterogeneity in Particulate Matter's Impact on Students' Scores Across Portuguese Scores' Distribution



Note: Plot of coefficients from four regressions analysing heterogeneity across the Portuguese score distribution. The x axis presents the coefficient of $PM_{2.5}$ per each gestational trimester. Each group of symbols represents the percentile explored as in the legend. The details on the specifications are the same as in table 4.3.

Figure 4.2: Heterogeneity in Particulate Matter's Impact on Students' Scores Across Maths Scores' Distribution



Note: Plot of coefficients from four regressions analysing heterogeneity across the Maths score distribution. The x axis presents the coefficient of $PM_{2.5}$ per each gestational trimester. Each group of symbols represents the percentile explored as in the legend. The details on the specifications are the same as in table 4.3.

4.4.4 School Performance and Wages

The literature on the returns of education has extensively showed that better cognitive test scores at school are positively related to individual higher earnings in the future (Murnane et al., 1995; Murphy and Peltzman, 2004; Blau and Kahn, 2005; Chetty et al., 2011; Currie et al., 2012; Curi and Menezes-Filho, 2015; Bharadwaj et al., 2017), to lower probability of drop-outs (Rivkin, 1995; Hanushek, 1996); and to higher economic growth (Hanushek and Kimko, 2000).

Curi and Menezes-Filho (2015) shows evidence of a 5% reduction of salary per 10% reduction of scores. Therefore, using our estimated cognitive performance effects and estimates on the returns to exams scores from Brazil (Curi and Menezes-Filho, 2015), we

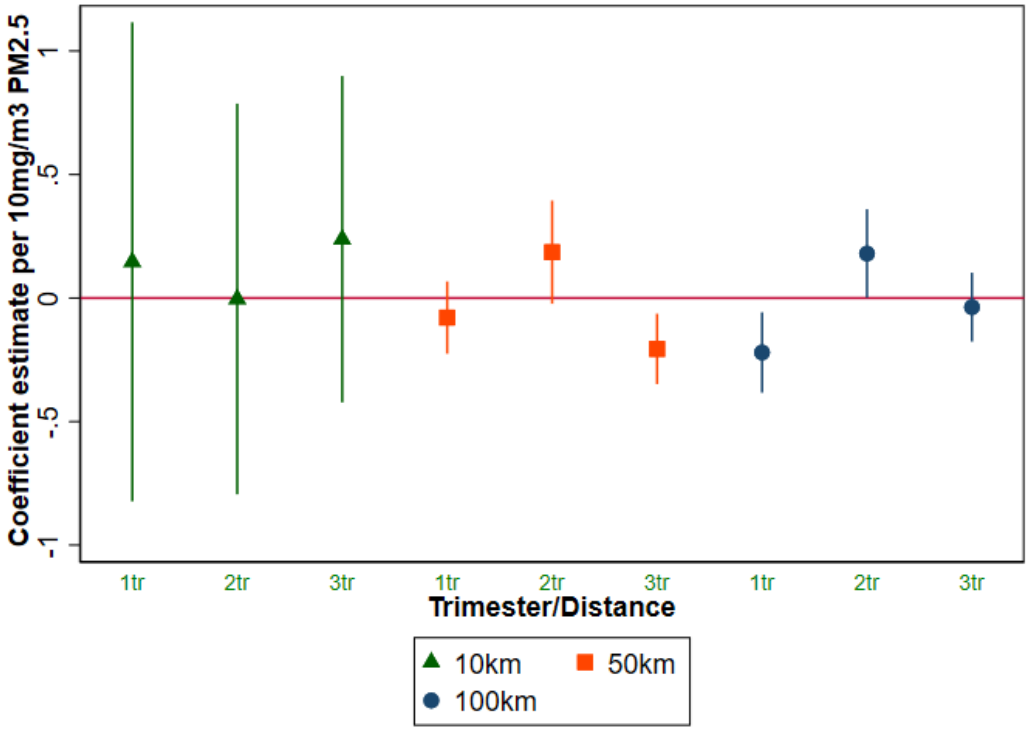
present a back-of-the-envelope calculation suggesting that a drop of $PM_{2.5}$ by 10% (*ceteris paribus*) would rise students' scores by 4.6%; and consequently elevate future individuals' earnings by 2.3%. Our calculations are based on Portuguese estimates. Overall, what we found is that being exposed to $10\mu g/m^3$ extra of $PM_{2.5}$ during the third trimester of pregnancy leads to a decrease in scores in the 5^{th} grade by 4.6% ((0.21 X 47.8)/219 = 4.6), where |0.21| is the coefficient from main estimation from Table 4.3, 47.8 and 219 are the SD and the average of Portuguese scores from Table 4.1). Bharadwaj et al. (2017) also found similar results when investigating fetal exposure to air pollution and students' scores in the fourth grade in Chile. The authors use the study on the returns to test scores in the United States by Blau and Kahn (2005) to provide their back-of-envelope calculation for Chilean students suggesting that, *ceteris paribus*, a reduction of 50% of Carbon Monoxide (CO) levels propels additional 1,000 US dollars in lifetime earnings per infant belonging to the cleaner cohort.

4.5 Sensitivity Analysis

To assess the internal validity of our research design and results we carry out a set of robustness tests. In the first exercise, we vary the distance between the fires' location and the municipalities' population centroid, using 10, 50 and 100 km (50km used in the main results). From Figures 4.3 and 4.4, one can observe that the effect of air pollution diminishes when the distance from the fires increases. This suggests that the pollutants' levels decay insofar as they travel, results consistent with Lai et al. (2018)'s findings.

Differently from the main specification in which the instruments used were the count of fires upwind and downwind, our second robustness exercise adds vertical fires as as instrument . The third category accounts for fires that are not upwind or downwind from the municipality's population centroid. Table 4.5 Columns 1-3 present results for the

Figure 4.3: Cognitive (Portuguese scores) impact of $PM_{2.5}$ per distance from the fires points to the population's municipality centroid

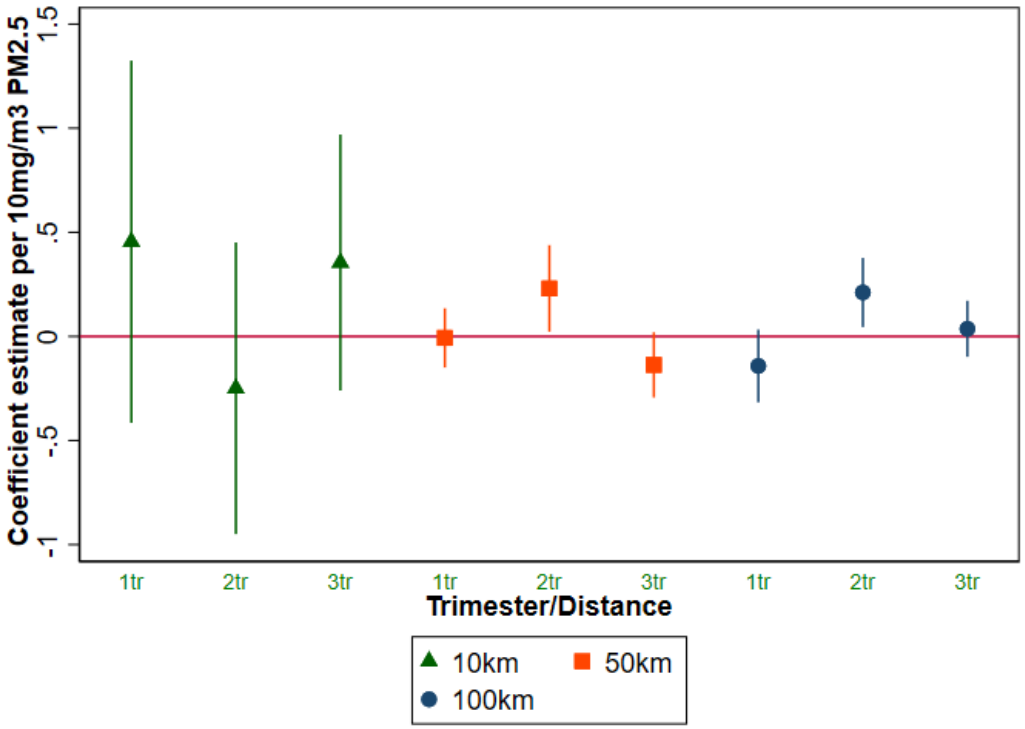


Note: This figure reports the cognitive (Portuguese scores) impact of air pollution within different distances. 10km is the baseline model. All the specifications follow the main model from table 4.3. We note an increase of standard errors suggesting that the further the stations the less precise and significant the estimates mainly for the third trimester.

outcome of Portuguese scores and Columns 4-6 for Maths. Columns 2 and 5 report the estimates for models with aggregated downwind and vertical fires in the same category as reported, with coefficients qualitatively unchanged when compared to the main specifications (Columns 1 and 4). In addition, in Columns 3 and 6, we utilise specifications with the three types of fires as instruments; i.e., upwind, downwind and vertical fires in the same equation. This latter test also shows that the main results (Column 1 and 4) remain qualitatively unchanged in terms of magnitude, signs, significance and the strength of instruments.

In our third set of robustness checks we slightly change the way we build our fire counts. In our baseline model, we utilise the confidence-weighted fire measure as the NASA's data set attributes a confidence index to each pixel-day as a measure of the prob-

Figure 4.4: Cognitive (Maths scores) impact of $PM_{2.5}$ per distance from the fires points to the population's municipality centroid



Note: This figure reports the cognitive (Maths scores) impact of air pollution within different distances. 10km is the baseline model. All the specifications follow the main model from table 4.3. We note an increase of standard errors suggesting that the further the stations the less precise and significant the estimates mainly for the third trimester.

ability that a fire occurred, ranging from 0% to 100%, based on weather conditions and vegetation at the exact moment the satellite captured the potential fire (Setzer and Sismanoglu, 2007; Rangel and Vogl, 2019). From Table 4.6 (Column 4 to 6), one can observe that the magnitude of the coefficients of the non-confidence weighted fires are somewhat smaller than the probability-weighted fires used in the main specification, with the First Stage F test still indicating strong instruments. Additionally, Table 4.7 suggests a slight shrinkage of the magnitude of air pollution's coefficients predicted utilising non-weighted fires (Columns 3 and 4).

To demonstrate that our baseline estimates with binned weather controls are not sensitive to different forms of meteorological controlling, we include a fourth robustness analysis. Table 4.8 presents our main specification results without accounting for weather

and with less flexible controls, i.e., with linear, squared, and interactions controls). The results remain generally the same, indicating that our main estimates are not driven by unobserved weather conditions correlated with wind and students' performance, but solely by the effects of air pollution.

In a fifth robustness exercise, Table 4.9 presents results of our estimations with SO_2 as co-pollutant ($PM_{2.5}$ and SO_2 present a pairwise correlation of 27.3% significant at 5% level). One can notice that the First Stage F statistics for both pollutants drops well below the threshold of 10, indicating that our fires instruments are not relevant to predict the latter pollutant. It is somehow a good sign suggesting that our baseline results using only $PM_{2.5}$ are not driven by other pollutants; that is, our instrumental variables approach helps us to disentangle the effect of $PM_{2.5}$ from other components of air pollution.

4.6 Conclusion

In this chapter, we examine the relationship between fetal exposure to air pollution from agricultural fires and cognitive performance on fifth grade exams in Brazil. To this end we merge data from the Brazilian Ministry of Education; NASA's satellite remote sensing data with active fires detection and air pollution data and atmospheric conditions gridded data from the Japanese Reanalysis archive, JRA-55. We use a set of controls for parents' level of education and children's characteristics to control for observable effects, as well as for municipalities unobserved fixed effects. Despite the inclusion of these control variables, some unobserved time-varying features can cause endogeneity in our analysis due to omission and classical measurement error inherent to this type of air pollution study. Thus, to address this problem we use an instrumental variable approach that we consider capable of creating exogenous shocks to our air pollution measure, $PM_{2.5}$.

Our empirical strategy, which relies upon changes in wind direction and fires' location as predictors of air pollution, yields credible evidence that the exposure to poor air quality during the gestational period causes a reduction in cognitive ability of students. More precisely, we argue that students living in municipalities located upwind from the fires present lower academic performance than their downwind dwellers counterparts. To put our findings into perspective, we demonstrate that students exposed to additional $10\mu g/m^3$ of $PM_{2.5}$ during the third trimester of pregnancy suffer a reduction of 4.6% in their Portuguese scores and 2.8% in Maths. On top of that, when we study the response from different sub-populations within our sample, we find that boys are more affected than girls. For example, males' literacy skills drop by 0.29SD, whereas females' Portuguese scores are reduced by 0.12SD per extra $10\mu g/m^3$ of $PM_{2.5}$ during the third gestational period. When it comes to numerical ability, boys' Maths scores fall by 0.19SD per $10\mu g/m^3$ of that pollutant, while girls' seem not to be significantly affected. These findings are consistent with previous studies on the impact of poor air quality and cognitive impairment including Jayachandran (2009), Sanders (2012), Ebenstein et al. (2016) and the third chapter of this thesis. Although the role of gender in the effects of air pollution on respiratory health remains unclear, some epidemiology studies point out that males are more affected than females. For example, Abbey et al. (1998) link PM_{10} exposure to reduced lung function among males but not females; and Galizia and Kinney (1999) find similar results regarding the exposure to O_3. Similarly, Ebenstein et al. (2016) find the effect of pollution on exam performance to differ by gender and cite the increased incidence of asthma amongst boys as a possible cause. Another interesting result emerged when exploring the scores' distribution. We find that very weak students (5^{th} percentile) and very top students (95^{th} percentile) are more sensitive to exposure during the third trimester of pregnancy than other pupils for both subjects. These results are in line with previous works, for example, Ebenstein et al. (2016) find evidence of students from the bottom of the distribution being more negatively impacted by bad air quality during the examinations, whereas, in Graff-Zivin et al. (2020)'s study, stronger pupils have their

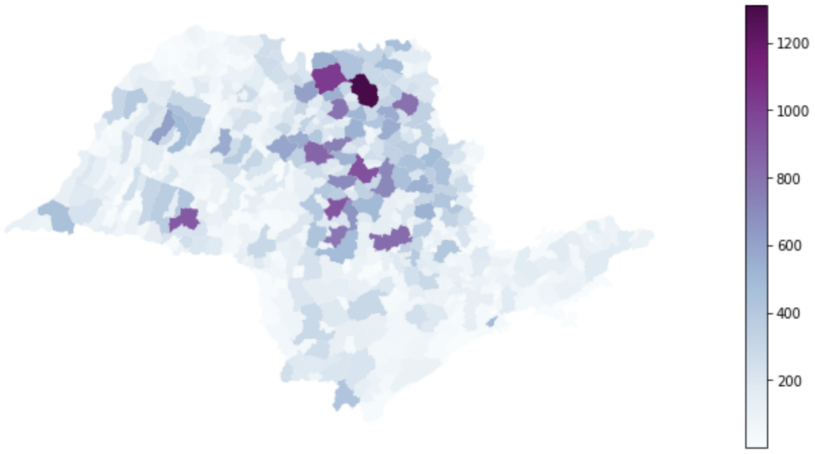
scores worsened by the level of pollution in the day of the tests.

To check the internal validity of our findings we perform a series of tests using different sets of instruments and distances between fires and municipality's population centroid, equations with co-pollutant SO_2 and falsification tests to which our results are robust and consistently suggest the harmful effect of fetal exposure to air pollution and cognitive health. In addition, we perform back-of-the-envelope calculations and show that a reduction of 10% of $PM_{2.5}$ during the third trimester of pregnancy has the potential to increase later life wages by 2.3%. These findings shed light on the important issue of wealth accumulation mainly in developing countries with sharper levels of inequality. Suffice it to say that the less well-off, who live in areas with low environmental standards, end up suffering more from the negative impacts of fetal exposure to air pollution on their cognitive performance, and hence deepening the wage inequality in Brazil.

All in all, our results seem to add plausible evidence suggesting that cognitive performance may be hindered by in utero exposure to bad air quality. As such, this chapter's findings may be useful to cast light on the imminent necessity of withdrawing traditional agricultural burning techniques, which, in fact, is due to occur soon. In fact, the government of the state of São Paulo has signed the Green Ethanol Protocol in 2017 due to start its validity in 2021 (SIMA, 2017) aimed at boosting the mechanisation of the harvesting phase as well as restoring the riparian areas. Although the fires have not started to decline yet, they seem to have plateaued (Aguiar et al., 2011), which is a positive sign towards the reduction of air pollution and, as demonstrated in our findings, it may entail long-run gains in economic growth through the human capital channel. In the short term, future research can be conducted after the implementation of this policy to assess its impacts on birth outcomes.

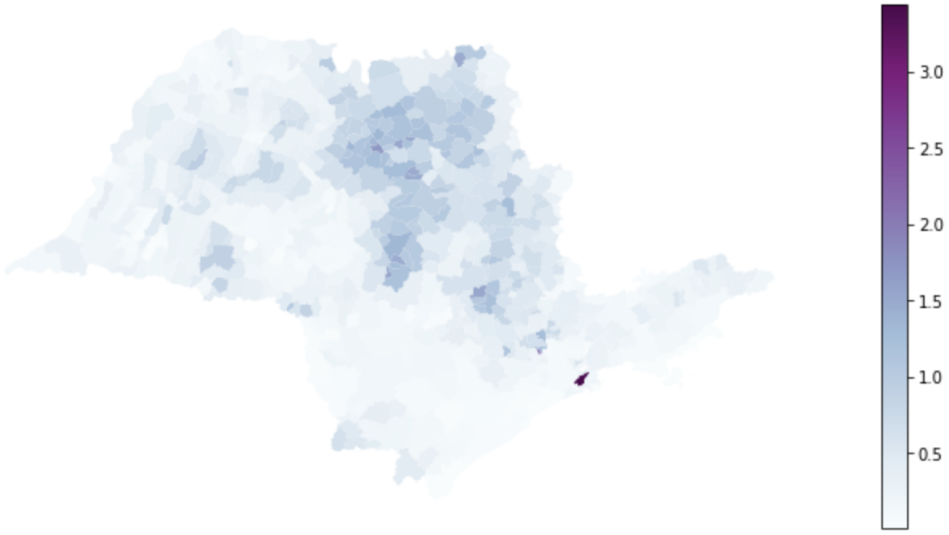
Appendix

Figure C1: Total of agricultural fires per municipality from 2000 to 2008



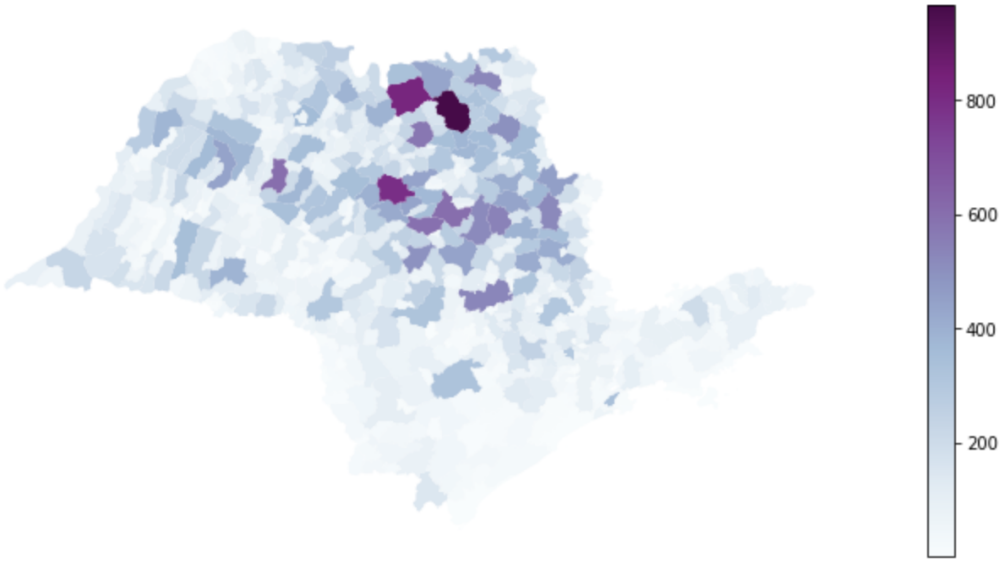
Note: Total amount of fires per municipality in the state of São Paulo during the birth years of our sample cohorts.

Figure C2: Total of agricultural fires per municipality size from 2000 to 2008



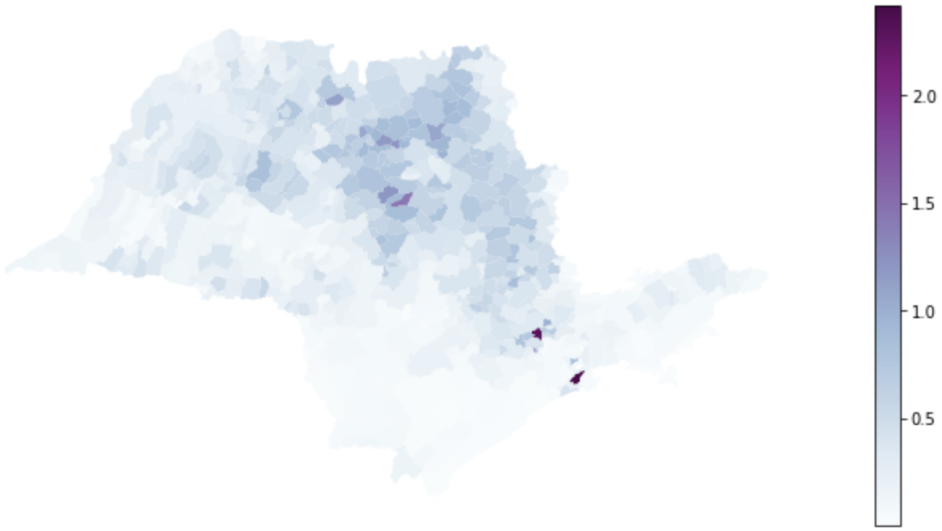
Note: Total amount of fires per square kilometers within each municipality of the state of São Paulo during the birth years of our sample cohorts.

Figure C3: Total of agricultural fires per municipality from 2009 to 2017



Note: Total amount of fires per municipality in the state of São Paulo in our samples' exams period.

Figure C4: Total of agricultural fires per municipality size from 2009 to 2017



Note: Total amount of fires per square kilometers within each municipality of the state of São Paulo in our samples' exams period.

Table 4.2: Instrument for Pollution: Count of fires per Wind Direction – First Stage Results)

Fires counts by wind direction	(1) PM_{25}_1tr	(2) PM_{25}_2tr	(3) PM_{25}_3tr	(4) PM_{25}_1tr	(5) PM_{25}_2tr	(6) PM_{25}_3tr	(7) PM_{25}_1tr	(8) PM_{25}_2tr	(9) PM_{25}_3tr
Upwind_1tr	0.00	-0.05**	-0.06**	-0.10	-0.01	-0.04	-0.09	-0.03	-0.02
	(0.05)	(0.02)	(0.03)	(0.06)	(0.02)	(0.05)	(0.06)	(0.02)	(0.06)
Downwind_1tr	0.31***	-0.14***	0.04**				0.20***	-0.12**	0.10**
	(0.07)	(0.04)	(0.02)				(0.06)	(0.06)	(0.05)
Upwind_2tr	0.05**	0.04	-0.06***	0.04*	-0.06	-0.04**	0.06***	-0.06	-0.04
	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.03)
Downwind_2tr	0.07**	0.25***	-0.04				0.06**	0.11***	-0.01
	(0.03)	(0.04)	(0.03)				(0.03)	(0.03)	(0.03)
Upwind_3tr	-0.10***	0.11***	0.06**	-0.08**	0.10***	-0.02	-0.06**	0.11***	-0.01
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
Downwind_3trsh	0.02	0.04*	0.22***				0.07*	0.01	0.15***
	(0.02)	(0.02)	(0.03)				(0.04)	(0.02)	(0.03)
Down+Vertical wind_1tr				0.15***	-0.06***	0.00			
				(0.03)	(0.01)	(0.02)			
Down+Vertical wind_2tr				0.02**	0.13***	-0.02**			
				(0.01)	(0.02)	(0.01)			
Down+Vertical wind_3tr				-0.00	0.01	0.11***			
				(0.01)	(0.01)	(0.01)			
Vertical wind_1tr							0.12***	-0.02	-0.05
							(0.04)	(0.02)	(0.05)
Vertical wind_2tr							-0.00	0.14***	-0.03
							(0.01)	(0.04)	(0.02)
Vertical wind_3tr							-0.05*	0.01	0.08***
							(0.03)	(0.02)	(0.02)
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.43	0.36	0.32	0.44	0.37	0.32	0.44	0.37	0.33

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured in 10 $\mu g/m^3$. Up, down and vertical wind represent the share of fires (weighted by confidence) per trimester of pregnancy by wind direction. Columns 1 to 3 present the first stage for the main specification with instruments upwind and downwind fires. Columns 4 to 6 present models using two winds categories: upwind; downwind + vertical wind). Columns 7 to 9 refer to models using three categories of wind: upwind, downwind and vertical wind fires confidence weighted.

Table 4.3: The Effects of $PM_{2.5}$ on Students' Scores - OLS and IV (Main Specification with flexible weather controls)

	Without Instruments		With Instruments	
Variables	(1)	(2)	(3)	(4)
	Panel A: Dep. Var. Portuguese Score (SD)			
$PM_{2.5}$_1tr	0.04**	0.03**	-0.08	-0.08
	(0.01)	(0.01)	(0.09)	(0.09)
$PM_{2.5}$_2tr	0.17***	0.15***	0.21**	0.18**
	(0.03)	(0.03)	(0.09)	(0.08)
$PM_{2.5}$_3tr	-0.03	-0.02	-0.24***	-0.21***
	(0.05)	(0.04)	(0.07)	(0.07)
Students' Controls	N	Y	N	Y
First stage F-test			14.9	14.9
Observations	1,190,717	1,187,660	1,190,717	1,187,660
R-squared	0.04	0.09	0.00	0.05
	Without Instruments		With Instruments	
Variables	(1)	(2)	(3)	(4)
	Panel B: Dep. Var. Mathematics Score (SD)			
$PM_{2.5}$_1tr	0.03	0.03*	-0.01	-0.01
	(0.02)	(0.02)	(0.08)	(0.08)
$PM_{2.5}$_2tr	0.16***	0.15***	0.24***	0.23***
	(0.02)	(0.02)	(0.09)	(0.08)
$PM_{2.5}$_3tr	-0.03	-0.03	-0.16**	-0.14*
	(0.04)	(0.04)	(0.08)	(0.08)
Students' Controls	N	Y	N	Y
First stage F-test			14.9	14.9
Observations	1,190,717	1,187,660	1,190,717	1,187,660
R-squared	0.06	0.10	0.00	0.04

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured in 10 $\mu g/m^3$. Instruments used upwind and downwind fires confidence weighted.

Table 4.4: Heterogeneity in Particulate Matter's Impact on Students' Scores per Gender

	Portuguese Scores (SD)			Mathematics Scores (SD)		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	All	Girls	Boys	All	Girls	Boys
PM_{25}_1tr	-0.08	-0.16*	-0.03	-0.01	-0.07	0.03
	(0.09)	(0.09)	(0.10)	(0.08)	(0.09)	(0.09)
PM_{25}_2tr	0.18**	0.09	0.25***	0.23***	0.14	0.29***
	(0.08)	(0.10)	(0.07)	(0.08)	(0.09)	(0.08)
PM_{25}_3tr	-0.21***	-0.12*	-0.29***	-0.14*	-0.05	-0.19**
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
First stage F-test	14.9	13.9	14.4	14.9	13.9	14.4
Observations	1,187,660	543,030	644,630	1,187,660	543,030	644,630
R-squared	0.05	0.05	0.04	0.04	0.04	0.04

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Instruments used upwind and downwind fires confidence weighted.

Table 4.5: The Effects of $PM_{2.5}$ on Students' Scores (using IV with different wind directions specifications)

	Portuguese Score (SD)			Mathematics Score (SD)		
	Main Model	Two Winds	Three Winds	Main Model	Two Winds	Three Winds
Variables	(1)	(2)	(3)	(4)	(5)	(6)
$PM_{2.5}_1tr$	-0.08	-0.10	-0.08	-0.01	-0.03	-0.01
	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)
$PM_{2.5}_2tr$	0.18**	0.15***	0.16***	0.23***	0.19***	0.20***
	(0.08)	(0.06)	(0.06)	(0.08)	(0.05)	(0.05)
$PM_{2.5}_3tr$	-0.21***	-0.23***	-0.22***	-0.14*	-0.15*	-0.14*
	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)
First stage F-test	14.9	25	16.1	14.9	25	16.1
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.05	0.05	0.05	0.04	0.04	0.04

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured in 10 $\mu g/m^3$. Column 1 presents the main specification with instruments upwind and downwind fires. Columns 2 and 5 present models using two winds categories: upwind; downwind + vertical wind). Columns 4 and 6 refer to models using three categories of wind: upwind, downwind and vertical wind fires confidence weighted.

Table 4.6: First Stage Results: Instruments for Pollution using non-confidence weighted fires

	Confidence Weighted Fires			Non Confidence Weighted Fires		
Fires counts by	(1)	(2)	(3)	(4)	(5)	(6)
wind direction	PM_{25}_1tr	PM_{25}_2tr	PM_{25}_3tr	PM_{25}_1tr	PM_{25}_2tr	PM_{25}_3tr
Upwind_1tr	0.00	-0.05**	-0.06**			
	(0.05)	(0.02)	(0.03)			
Downwind_1tr	0.31***	-0.14***	0.04**			
	(0.07)	(0.04)	(0.02)			
Upwind_2tr	0.05**	0.04	-0.06***			
	(0.02)	(0.03)	(0.02)			
Downwind_2tr	0.07**	0.25***	-0.04			
	(0.03)	(0.04)	(0.03)			
Upwind_3tr	-0.10***	0.11***	0.06**			
	(0.03)	(0.02)	(0.02)			
Downwind_3tr	0.02	0.04*	0.22***			
	(0.02)	(0.02)	(0.03)			
Upwind_1tr				-0.03	-0.03**	-0.05***
				(0.04)	(0.01)	(0.02)
Downwind_1tr				0.24***	-0.11***	0.04***
				(0.05)	(0.03)	(0.01)
Upwind_2tr				0.02	0.00	-0.04***
				(0.02)	(0.02)	(0.01)
Downwind_2tr				0.07***	0.18***	-0.04*
				(0.03)	(0.02)	(0.02)
Upwind_3tr				-0.07***	0.08***	0.02
				(0.02)	(0.01)	(0.02)
Downwind_3tr				0.01	0.04***	0.18***
				(0.02)	(0.01)	(0.02)
First stage F-test	14.9	14.9	14.9	14.8	14.8	14.8
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660

Notes: Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Columns 1 to 3 report results using upwind and downwind fires weighted by the probability of occurrence; Columns 4 to 6 report results using upwind and downwind fires non-weighted by the probability of occurrence.

Table 4.7: Particulate Matter's Impact on Students' Scores using fires non-confidence weighted

	Portuguese Scores (SD)		Mathematics Scores (SD)	
	(1)	(2)	(3)	(4)
	Confidence	Non-confidence	Confidence	Non-confidence
Variables	Weighted Fires	Weighted Fires	Weighted Fires	Weighted Fires
PM_{25}_1tr	-0.08	-0.09	-0.01	-0.03
	(0.09)	(0.08)	(0.08)	(0.07)
PM_{25}_2tr	0.18**	0.21**	0.23***	0.25***
	(0.08)	(0.09)	(0.08)	(0.08)
PM_{25}_3tr	-0.21***	-0.21***	-0.14*	-0.15*
	(0.07)	(0.07)	(0.08)	(0.08)
First stage F-test	14.9	14.8	14.9	14.8
Observations	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.05	0.05	0.04	0.04

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Instruments used upwind and downwind fires non-confidence weighted.

Table 4.8: The Effects of $PM_{2.5}$ on Students' Scores (SD) (without and with linear weather controls)

	Without IV		With IV		Without IV		With IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Portuguese Scores (SD)				Mathematics Scores (SD)			
PM_{25}_1tr	-0.01	-0.10	0.03*	-0.06	-0.02	-0.05	0.02	0.01
	(0.01)	(0.06)	(0.01)	(0.09)	(0.01)	(0.06)	(0.02)	(0.08)
PM_{25}_2tr	0.12***	0.13	0.15***	0.22**	0.11***	0.17**	0.14***	0.26***
	(0.02)	(0.08)	(0.02)	(0.09)	(0.02)	(0.08)	(0.02)	(0.09)
PM_{25}_3tr	-0.02	-0.07	-0.02	-0.19***	-0.02	0.01	-0.02	-0.12
	(0.03)	(0.08)	(0.04)	(0.07)	(0.03)	(0.08)	(0.04)	(0.08)
Weather Controls	N	N	Y	Y	N	N	Y	Y
First stage F-test			29.2	16			29.2	16
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.09	0.05	0.09	0.05	0.09	0.04	0.10	0.04

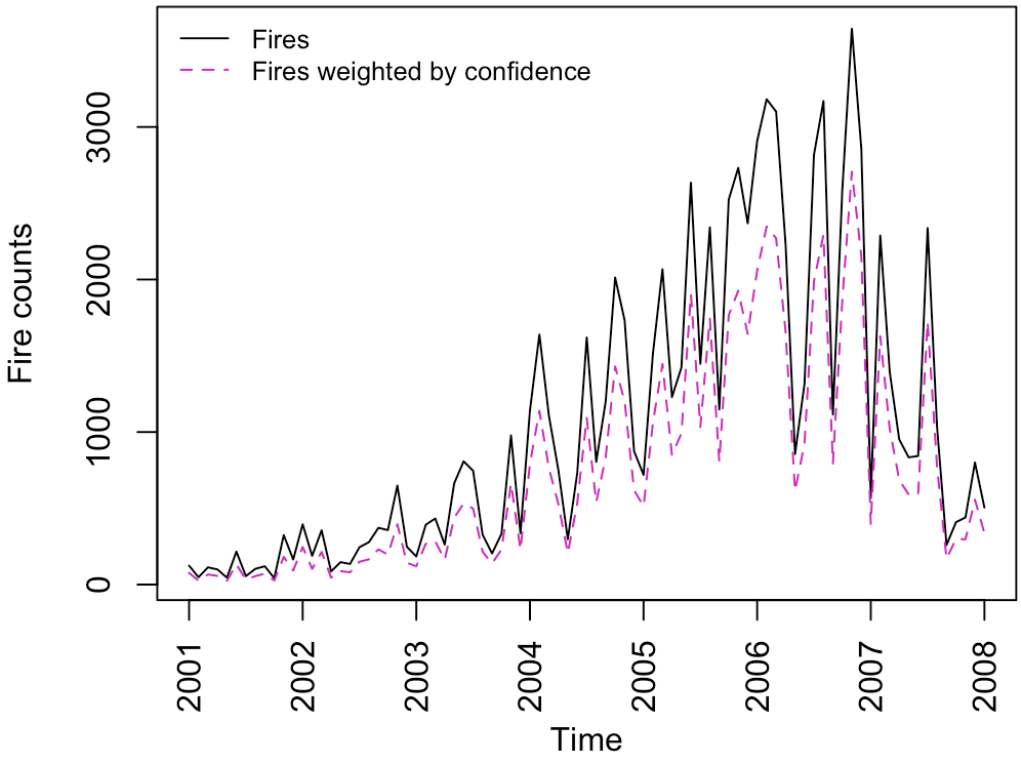
Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature linear, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Instruments used upwind and downwind fires confidence weighted.

Table 4.9: The Effects of $PM_{2.5}$ on Students' Scores (SD) (with co-pollutant)

	Without IV		With IV		Without IV		With IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Portuguese Scores (SD)				Mathematics Scores (SD)			
PM_{25}_1tr	0.03**	0.04***	-0.08	-0.08	0.03*	0.04***	-0.01	-0.01
	(0.01)	(0.02)	(0.09)	(0.08)	(0.02)	(0.02)	(0.08)	(0.08)
PM_{25}_2tr	0.15***	0.18***	0.18**	0.19**	0.15***	0.19***	0.23***	0.24***
	(0.03)	(0.03)	(0.08)	(0.08)	(0.02)	(0.03)	(0.08)	(0.08)
PM_{25}_3tr	-0.02	-0.01	-0.21***	-0.21***	-0.03	-0.01	-0.14*	-0.15*
	(0.04)	(0.04)	(0.07)	(0.07)	(0.04)	(0.03)	(0.08)	(0.08)
SO_2_1tr		0.02		0.00		0.00		0.00
		(0.04)		(0.00)		(0.04)		(0.00)
SO_2_2tr		-0.12***		0.00		-0.15***		0.00
		(0.04)		(0.00)		(0.05)		(0.00)
SO_2_3tr		-0.11*		0.00		-0.11		0.00
		(0.06)		(0.00)		(0.07)		(0.00)
First stage F-test			14.9	0.35			14.9	0.35
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.09	0.09	0.05	0.05	0.10	0.10	0.04	0.04

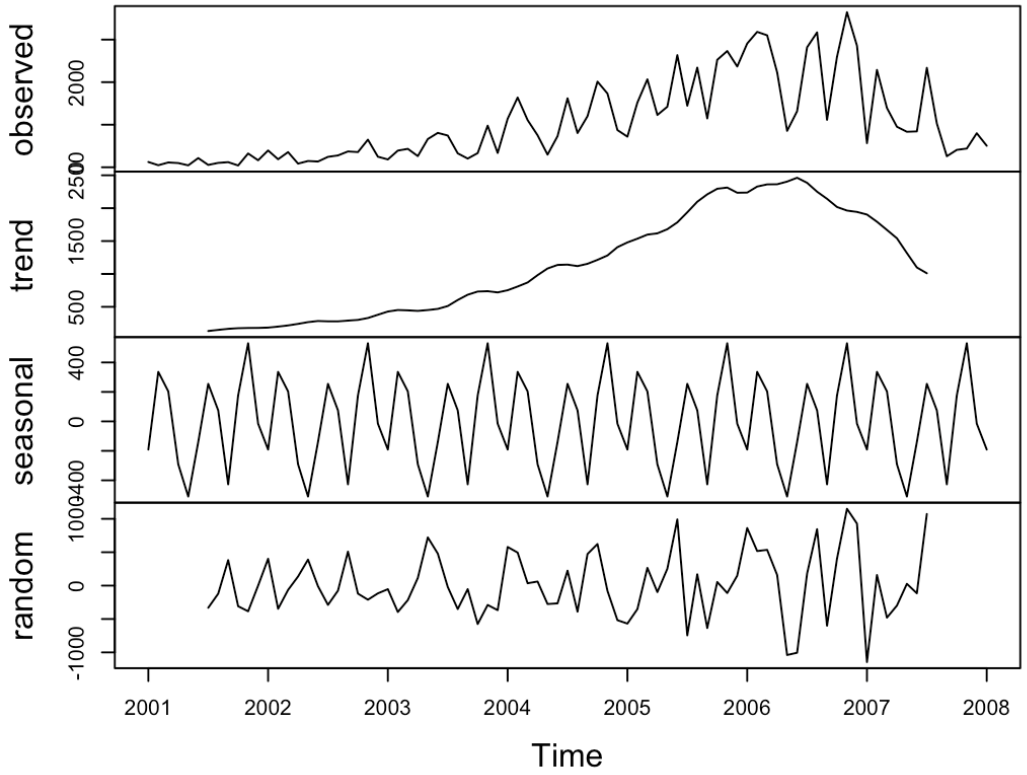
Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr, PM_{25}_3tr, SO_2_1tr, SO_2_2tr and SO_2_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Instruments used upwind and downwind fires confidence weighted.

Figure C5: Monthly average of counts of fires from 2001-2008



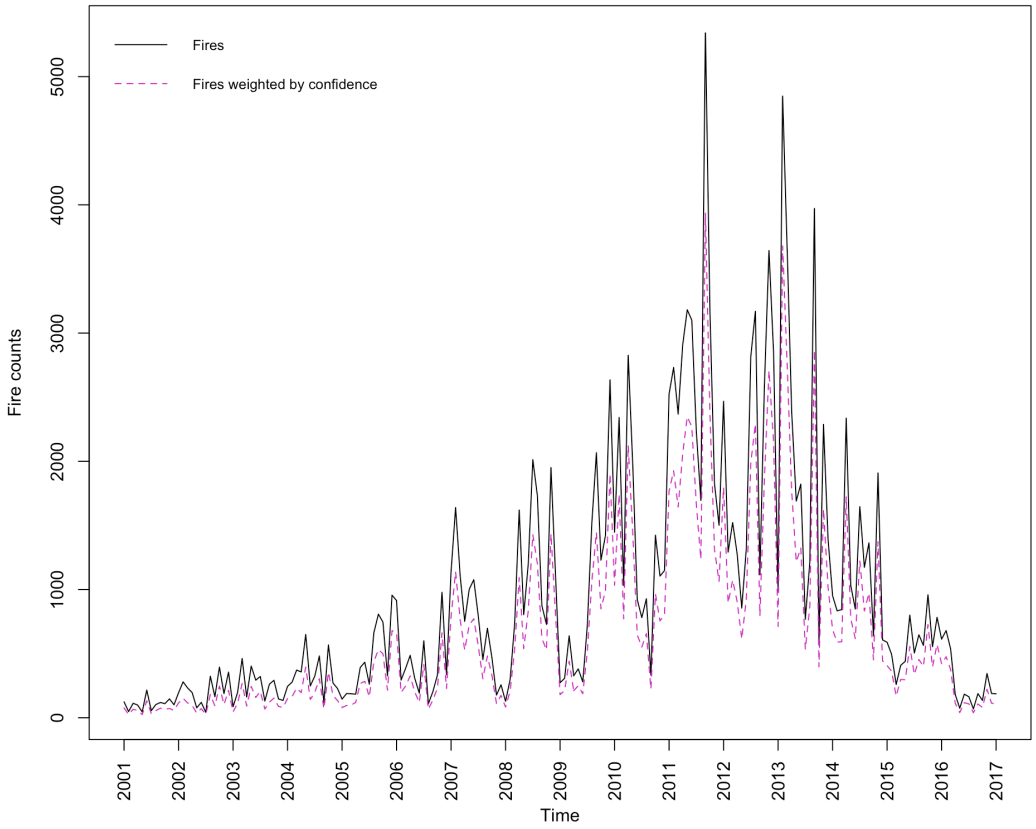
Note: The graph shows the monthly average of counts of fires and fires confidence-weighted from 2001-2008 in São Paulo state. We can see a that the peak of fires occur during the winter months.

Figure C6: Decomposition Additive Time Series of monthly average counts of fires from 2001-2008



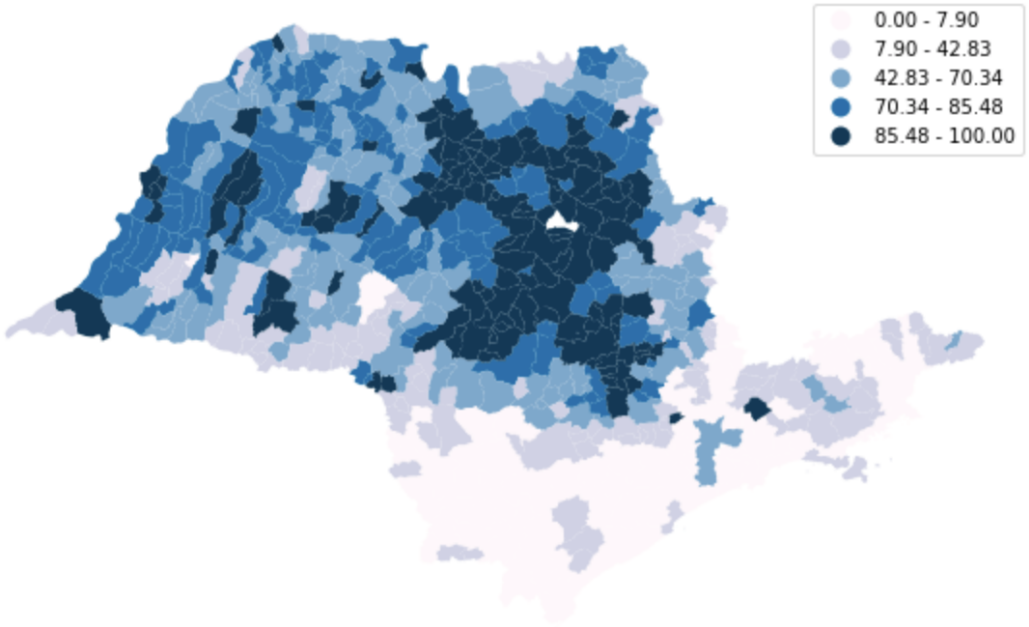
Note: The graphs shows Decomposition Additive Time Series of monthly average counts of fires from 2001-2008 in São Paulo state. We can see a strong seasonality with the peak of fires occur during the winter months and a not clear time trend.

Figure C7: Monthly average of counts of fires from 2001-2017



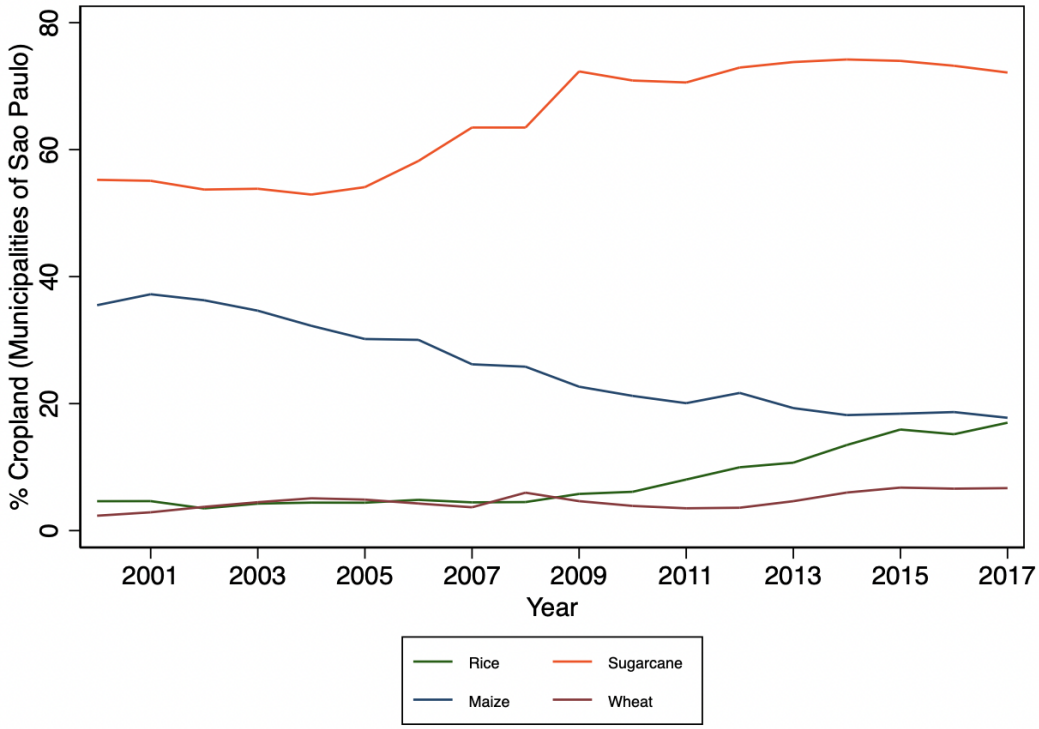
Note: The graph shows the monthly average of counts of fires and fires confidence-weighted from 2001-2017 in São Paulo state. We can see a that the peak of fires occur during the winter months.

Figure C8: Share of sugarcane crop per municipality



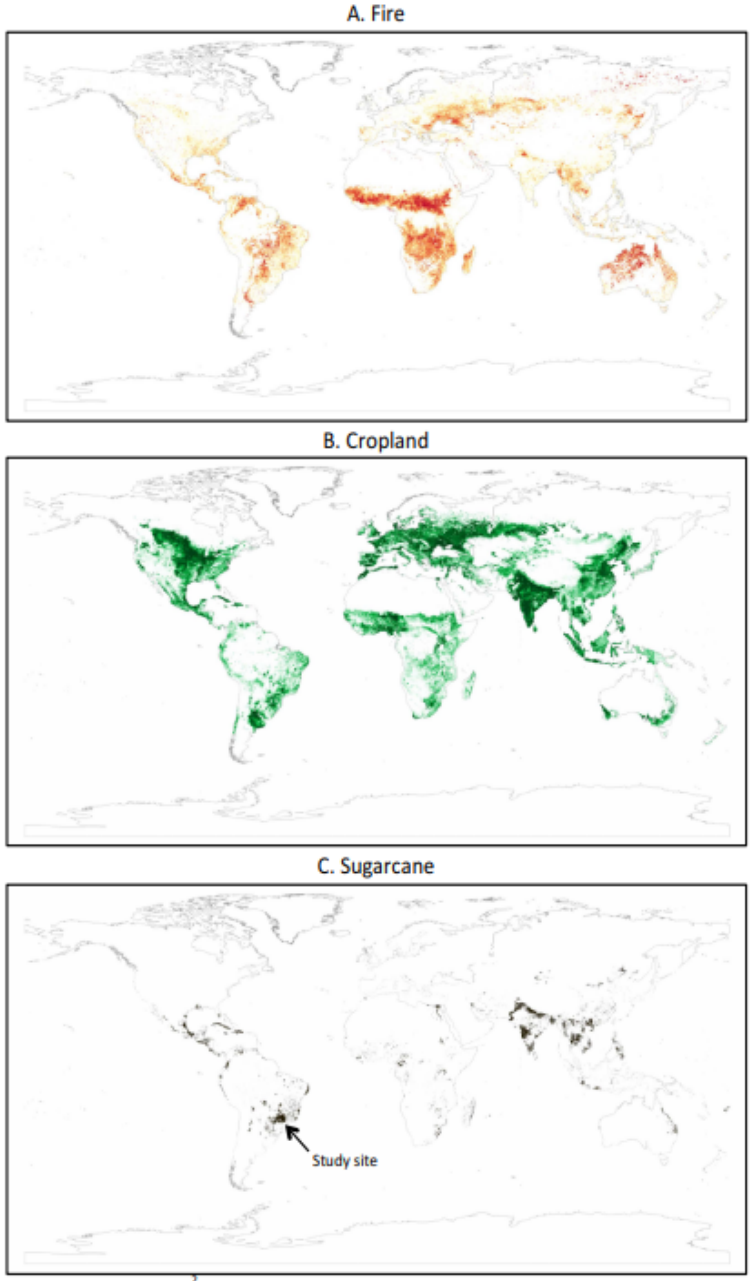
Note:This figure shows the share of sugarcane crop per municipality from 2000 to 2017. One can notice that the areas with higher concentration of this crop coincide with the areas with largest number of fire counts (C1 and C3).

Figure C9: Share of rice, sugarcane, maize and wheat planted per municipality from 2000 to 2017



Note:This figure depicts the proportion of land destined to the plantation of four crops that potentially utilise agricultural burning. One can note that sugarcane is the crop that uses this technique the most in the State of São Paulo throughout the period.

Figure C10: Distribution of Fires, cropland and sugarcane in early 2000s



Note: 5 arcminute cells, approx. 100km^2 at Equator. Darker areas indicate more fire, cropland, or sugarcane. Panel A is based on remote sensing fires data from the TERRA satellite for Nov. 2000 – Oct. 2001. Panel B is from Ramankutty et al. (2008), who estimate cropland cover circa 2000 by combining satellite remote sensing and agricultural inventory data. Panel C is from You et al. (2017), who estimate sugarcane harvest area circa 2005 using similar methods.

We base our analysis on Ramankutty, Navin, A.T. Evan, C. Monfreda, and J.A. Foley. (2008). "Farming the Planet: Geographic Distribution of Global Agricultural Lands in the Year 2000." Global Biogeochemical Cycles 22(1); and You, L., U. Wood-Sichra, S. Fritz, Z. Guo, L. See, and J. Koo. (2017). Spatial Production Allocation Model (SPAM) 2005 v3.2.

Source: Rangel and Vogl (2019). Note: This figure depicts the global distribution of fires, cropland and sugarcane. Fires relate to the period from Nov.2000 - Oct.2001. Cropland, year 2000. Sugarcane, year 2005.

Chapter 5

Conclusion

This thesis consists of three independent studies on human capital in Brazil. More precisely, I investigate the impact of educational policies, air pollution and agricultural fires on students' cognitive performance. In the first chapter, I present an outline for this thesis. In Chapter Two, I explore two educational policies as treatment variables and their impact on elementary school pupils' performance. In Chapter Three, I study the impact of air pollution in the day of the nationwide university entrance examinations on the candidates' scores. Finally, in Chapter Four, I continue to investigate the harmful effects of fetal exposure to bad air quality on fifth graders' cognitive performance, but stemming from agricultural fires from sugarcane harvesting phase.

In Chapter Two, I analyse the effect of two independent educational policies on students' performance at the national examination *Prova Brasil*, dropout and pass rates. The first policy made it optional for schools to adopt the automatic promotion from the first to the ninth grades, and the second policy increased the compulsory years of schooling from 8 to 9 years, changing the starting age from 7 to 6 years old. Regarding the first policy, I found evidence evidence suggesting that it indeed increase the pass rates as expected. However, the results indicate also a positive relationship between non-repetition in the primary school and dropout rates, which may be signalling that the policy itself

does not suffice to change the alarming dropout scenario in Brazil. With regards to the second policy which increased the compulsory years of schooling from 8 to 9 years, my findings indicate that this policy raised the Portuguese and Maths' scores in the national examinations, but also increased the dropout rates and lowered the pass rates.

These results suggest that it is not only by issuing new educational rules that the problem of low quality of education in Brazil will be sorted. For example, new staff must be trained, monitoring should apply not only the students, but also the schools, and the professionals. These remarks can be applied to other developing countries as well, with governments ensuring adequate availability of complementary tools to influence the accomplishment of the policies objectives, the efficiency and quality of primary education. Future research can cast light on the compulsory SSA policy phase, when there is data set available to assess the causality of the reform on the response variables analysed in this paper.

In Chapter Three, I investigate the contemporaneous effect of air pollution on students' scores at university entrance examination. I find evidence that bad air quality in the days of the exams hinders cognitive performance, with boys and poorer students more affected. In Chapter Four, I turn my attention to air pollution from sugarcane fires during pregnancy. My results for this research suggest that the fetal exposure to fires' smoke causes impairments in cognitive health later in life.

In Chapter Four, I explore agricultural fires from sugarcane crops in the State of São Paulo as air pollution sources. More specifically, I look at the effects of the fetal exposure to bad air quality stemming from these fires on fifth graders scores and wages later in life. Once more, I show evidence of the harmful effects of air pollution on cognitive performance, this time in the medium and long run.

These two studies shed light on the important issue of air pollution, mainly in developing countries where environmental regulations and quality of education are weaker than

in developed nations. Thus, my research topics contribute to the understanding of the effects of bad air quality in the largest developing country in the Americas. Nonetheless, as in any social sciences research, my works present a few limitations in terms of data. As for the first chapter, had we had data for the compulsory phase of the education policy that increased the years of schooling, we would have been able to identify the causal relationship between that policy and students' performance in a more accurate way. However, given that the policy is recent, there is simply no time span to assess subsequent cohorts impacted by the mandatory phase. Regarding the third and fourth chapters, a common limitation is the lack of information on the students' residence and exact place of birth, which may bring some measurement error to our analysis since we use pollution data from the municipality's population centroid as a proxy for the actual address. Yet, this is a very recurrent issue that the literature on air pollution faces. As per the fourth chapter, it would be better if we had the infants' exact date of birth. This would enable us to precisely calculate the pregnancy trimesters and, hence, reduce attenuation bias concerns. Finally, future research can extend this study by looking at the compulsory phase of the educational policies from chapter 2, and wild fires from forest burning on health and other human capital outcomes *vis-à-vis* chapter 4.

Bibliography

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19.

Abadie, A., Drukker, D., Herr, J. L., Imbens, G. W., et al. (2004). Implementing matching estimators for average treatment effects in stata. *Stata journal*, 4:290–311.

Abbey, D. E., Burchette, R. J., Knutsen, S. F., McDonnell, W. F., Lebowitz, M. D., and Enright, P. L. (1998). Long-term particulate and other air pollutants and lung function in nonsmokers. *American journal of respiratory and critical care medicine*, 158(1):289–298.

Aguiar, D. A., Rudorff, B. F. T., Silva, W. F., Adami, M., and Mello, M. P. (2011). Remote sensing images in support of environmental protocol: Monitoring the sugarcane harvest in são paulo state, brazil. *Remote Sensing*, 3(12):2682–2703.

Akagi, S., Yokelson, R. J., Wiedinmyer, C., Alvarado, M., Reid, J., Karl, T., Crounse, J., and Wennberg, P. (2011). Emission factors for open and domestic biomass burning for use in atmospheric models. *Atmospheric Chemistry and Physics*, 11(9):4039–4072.

Akerlof, G. A. and Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of economic literature*, 40(4):1167–1201.

Almond, D., Currie, J., and Duque, V. (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature*, 56(4):1360–1446.

Almond, D. and Edlund, L. (2009). Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in sweden. *The Quarterly journal of economics*, 124(4):1729–1772.

Anderson, M. L. (2015). As the wind blows: The effects of long-term exposure to air pollution on mortality. *NBER Working Paper*, (w21578).

Andreae, M. O. and Merlet, P. (2001). Emission of trace gases and aerosols from biomass burning. *Global biogeochemical cycles*, 15(4):955–966.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arbex, M. A., Cançado, J. E. D., Pereira, L. A. A., Braga, A. L. F., and Saldiva, P. H. d. N. (2004). Biomass burning and its effects on health. *J Bras Pneumol*, 30(2):158–75.

Arbex, M. A., Martins, L. C., de Oliveira, R. C., Pereira, L. A. A., Arbex, F. F., Cançado, J. E. D., Saldiva, P. H. N., and Braga, A. L. F. (2007). Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in brazil. *Journal of Epidemiology & Community Health*, 61(5):395–400.

Arceo, E., Hanna, R., and Oliva, P. (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from mexico city. *The Economic Journal*, 126(591):257–280.

Archsmith, J., Heyes, A., and Saberian, S. (2018). Air quality and error quantity: Pollution and performance in a high-skilled, quality-focused occupation. *Journal of the Association of Environmental and Resource Economists*, 5(4):827–863.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, 28(25):3083–3107.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, 122(3):1235–1264.

Banzhaf, H. S. and Walsh, R. P. (2008). Do people vote with their feet? an empirical test of tiebout. *American Economic Review*, 98(3):843–63.

Barros, R. P. (2015). Educacao e produtividade. https://www.bndes.gov.br/wps/portal/site/home/conhecimento/seminarios/seminarios-e-apresentacoes-itens/educacao-produtividade-resumo-apresentacao. [Online; accessed 15-November-2018].

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2):9–49.

Becker, S., Ichino, A., et al. (2002). Estimation of average treatment effects based on propensity scores. *The stata journal*, 2(4):358–377.

Bedi, A. S., Nakaguma, M. Y., Restrepo, B. J., and Rieger, M. (2020). Particle pollution and cognition: Evidence from sensitive cognitive tests in brazil. *Journal of the Association of Environmental and Resources Economists (forthcoming)*.

Bell, M. L., Peng, R. D., and Dominici, F. (2006). The exposure–response curve for ozone and risk of mortality and the adequacy of current ozone regulations. *Environmental health perspectives*, 114(4):532–536.

Berlinski, S., Galiani, S., and Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of public Economics*, 92(5-6):1416–1440.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Betthäuser, B. (2017). Fostering equality of opportunity? compulsory schooling reform and social mobility in germany. *European Sociological Review*, 33(5):633–644.

Bharadwaj, P., Eberhard, J., and Neilson, C. (2013). Health at birth, parental investments and academic outcomes. *University of California at San Diego Working Paper Series*, pages 1862–1891.

Bharadwaj, P., Eberhard, J. P., and Neilson, C. A. (2018). Health at birth, parental investments, and academic outcomes. *Journal of Labor Economics*, 36(2):349–394.

Bharadwaj, P., Gibson, M., Graff Zivin, J., and Neilson, C. (2017). Gray matters: Fetal pollution exposure and human capital formation. *Journal of the Association of Environmental and Resource Economists*, 4(2):505–542.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too young to leave the nest? the effects of school starting age. *The Review of Economics and Statistics*, 93(2):455–467.

Blau, F. D. and Kahn, L. M. (2005). Do cognitive test scores explain higher us wage inequality? *Review of Economics and statistics*, 87(1):184–193.

Blundell, R. and Costa Dias, M. (2000). Evaluation methods for non-experimental data. *Fiscal studies*, 21(4):427–468.

Bondy, M., Roth, S., and Sager, L. (2020). Crime is in the air: The contemporaneous relationship between air pollution and crime. *Journal of the Association of Environmental and Resource Economists*, 7(3):555–585.

Bruns, B., Evans, D., and Luque, J. (2011). *Achieving world-class education in Brazil: The next agenda*. The World Bank.

Buchard, V., Da Silva, A., Randles, C., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J., and Winker, D. (2016). Evaluation of the surface pm2. 5 in version 1 of the nasa merra aerosol reanalysis over the united states. *Atmospheric Environment*, 125:100–111.

Calderón-Garcidueñas, L., Engle, R., Mora-Tiscareño, A., Styner, M., Gómez-Garza, G., Zhu, H., Jewells, V., Torres-Jardón, R., Romero, L., Monroy-Acosta, M. E., et al. (2011). Exposure to severe urban air pollution influences cognitive outcomes, brain volume and systemic inflammation in clinically healthy children. *Brain and cognition*, 77(3):345–355.

Calderón-Garcidueñas, L., Mora-Tiscareño, A., Ontiveros, E., Gómez-Garza, G., Barragán-Mejía, G., Broadway, J., Chapman, S., Valencia-Salazar, G., Jewells, V., Maronpot, R. R., et al. (2008). Air pollution, cognitive deficits and brain abnormalities: a pilot study with children and dogs. *Brain and cognition*, 68(2):117–127.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1):31–72.

Cançado, J. E., Saldiva, P. H., Pereira, L. A., Lara, L. B., Artaxo, P., Martinelli, L. A., Arbex, M. A., Zanobetti, A., and Braga, A. L. (2006). The impact of sugar cane–burning emissions on the respiratory system of children and the elderly. *Environmental health perspectives*, 114(5):725–729.

Cao, J., Yang, C., Li, J., Chen, R., Chen, B., Gu, D., and Kan, H. (2011). Association between long-term exposure to outdoor air pollution and mortality in china: a cohort study. *Journal of hazardous materials*, 186(2-3):1594–1600.

Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics*, volume 3, pages 1801–1863. Elsevier.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.

Card, D. and Krueger, A. B. (1993). Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania. Technical report, National Bureau of Economic Research.

Carvalho, J. a. C. d. (2014). *Uma maior ameaça de reprovação faz os alunos estudarem mais? Uma análise do impacto da volta ao regime seriado nas escolas públicas de Ensino Fundamental: efeitos agregados e dinâmicos*. PhD thesis, University of São Paulo - Ribeirão Preto.

Casen (2006). Encuesta de caracterización socioeconómica nacional casen. `https://www.mideplan.cl/casen`. [Online; accessed 15-November-2018].

Chang, T., Graff Zivin, J., Gross, T., and Neidell, M. (2016). Particulate pollution and the productivity of pear packers. *American Economic Journal: Economic Policy*, 8(3):141–69.

Chang, T., Graff Zivin, J., Gross, T., and Neidell, M. (2019). The effect of pollution on worker productivity: evidence from call center workers in china. *American Economic Journal: Applied Economics*, 11(1):151–72.

Chay, K. Y. and Greenstone, M. (2003). The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *The quarterly journal of economics*, 118(3):1121–1167.

Chen, Huang, W., Wong, C.-M., Wang, Z., Thach, T. Q., Chen, B., Kan, H., Group, C. C., et al. (2012). Short-term exposure to sulfur dioxide and daily mortality in 17 chinese cities: the china air pollution and health effects study (capes). *Environmental research*, 118:101–106.

Chen, Li, C., Ristovski, Z., Milic, A., Gu, Y., Islam, M. S., Wang, S., Hao, J., Zhang, H., He, C., et al. (2017a). A review of biomass burning: Emissions and impacts on air quality, health and climate in china. *Science of the Total Environment*, 579:1000–1034.

Chen, S., Guo, C., and Huang, X. (2018). Air pollution, student health, and school absences: Evidence from china. *Journal of Environmental Economics and Management*, 92:465–497.

Chen, X., Zhang, X., and Zhang, X. (2017b). Smog in our brains: Gender differences in the impact of exposure to air pollution on cognitive performance.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly journal of economics*, 126(4):1593–1660.

Clougherty, J. E. (2010). A growing role for gender analysis in air pollution epidemiology. *Environmental health perspectives*, 118(2):167–176.

Cochran, W. G. and Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 417–446.

Crawford, C., Dearden, L., and Greaves, E. (2014). The drivers of month-of-birth differences in children's cognitive and non-cognitive skills. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177(4):829–860.

Crawford, C., Dearden, L., and Meghir, C. (2007). When you are born matters: The impact of date of birth on child cognitive outcomes in england.

Crawford, C., Dearden, L., and Meghir, C. (2010). When you are born matters: the impact of date of birth on educational outcomes in england. Technical report, IFS working papers.

Curi, A. Z. and Menezes-Filho, N. A. (2015). A relação entre o desempenho escolar e os salarios no brasil. Technical report, Fundação Itau Social.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47(1):87–122.

Currie, J., Graff Zivin, J., Mullins, J., and Neidell, M. (2014). What do we know about short-and long-term effects of early-life exposure to pollution? *Annu. Rev. Resour. Econ.*, 6(1):217–247.

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., and Rivkin, S. G. (2009). Does pollution increase school absences? *The Review of Economics and Statistics*, 91(4):682–694.

Currie, J. and Neidell, M. (2005). Air pollution and infant health: what can we learn from california's recent experience? *The Quarterly Journal of Economics*, 120(3):1003–1030.

Currie, J., Thomas, D., et al. (2012). Early test scores, school quality and ses: Longrun effects on wage and employment outcomes. *35th Anniversary Retrospective*, 35:185–214.

De Loecker, J. (2007). Do exports generate higher productivity? evidence from slovenia. *Journal of international economics*, 73(1):69–98.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., and Reif, J. (2019a). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, 109(12):4178–4219.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., and Reif, J. (2019b). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, 109(12):4178–4219.

Deschênes, O., Greenstone, M., and Guryan, J. (2009). Climate change and birth weight. *American Economic Review*, 99(2):211–17.

Developers, E. C. (2016). Health effects of ozone in the general population. *United States Enviromental Protection Agency*.

Dobkin, C. and Ferreira, F. (2010). Do school entry laws affect educational attainment and labor market outcomes? *Economics of education review*, 29(1):40–54.

Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris Jr, B. G., and Speizer, F. E. (1993). An association between air pollution and mortality in six us cities. *New England journal of medicine*, 329(24):1753–1759.

Drange, N., Havnes, T., and Sandsør, A. M. (2012). Kindergarten for all: Long run effects of a universal intervention. Technical report, Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit.

Ebenstein, A., Fan, M., Greenstone, M., He, G., and Zhou, M. (2017). New evidence on the impact of sustained exposure to air pollution on life expectancy from china's huai river policy. *Proceedings of the National Academy of Sciences*, 114(39):10384–10389.

Ebenstein, A., Lavy, V., and Roth, S. (2016). The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics*, 8(4):36–65.

Erickson, L. D., Gale, S. D., Anderson, J. E., Brown, B. L., and Hedges, D. W. (2020). Association between exposure to air pollution and total gray matter and total white matter volumes in adults: A cross-sectional study. *Brain Sciences*, 10(3):164.

Estadao (2019). Infográficos. https://www.estadao.com.br/infograficos/educacao,no-enem-1-a-cada-4-alunos-de-classe-media-triunfa-pobres-sao-1-a-cada-600,953041. [Online; accessed 12-November-2019].

Estes Jr, M. G., Al-Hamdan, M. Z., Crosson, W., Estes, S. M., Quattrochi, D., Kent, S., and McClure, L. A. (2010). What can affect aod–pm2. 5 association? *Environmental Health Perspectives*, 118(3).

Eyles, A. and Machin, S. J. (2015). The introduction of academy schools to england's education.

FAOSTAT (2020). Fao statistics. http://faostat.fao.org/FAOSTATFAOStatisticsDivision. [Online; accessed 19-August-2020].

Fernandes, A. (1988). A utilização da queimada na colheita da cana-de-açúcar. *COPERSUCAR CTAG-1*, pages 1–17.

Ferrão, M. E., Beltão, K. I., and dos Santos, D. P. (2002). Políticas de não-repetência e a qualidade da educação: evidências obtidas a partir da modelagem dos dados da 4ª série do saeb-99. *Estudos em avaliação educacional*, (26):47–74.

Fertig, M. and Kluve, J. (2005). The effect of age at school entry on educational attainment in germany. *RWI*, (27). Discussion Papers.

Galizia, A. and Kinney, P. L. (1999). Long-term residence in areas of high ozone: associations with respiratory health in a nationwide sample of nonsmoking young adults [dsee comments]. *Environmental health perspectives*, 107(8):675–679.

Gan, W. Q., Koehoorn, M., Davies, H. W., Demers, P. A., Tamburic, L., and Brauer, M. (2011). Long-term exposure to traffic-related air pollution and the risk of coronary heart disease hospitalization and mortality. *Environmental health perspectives*, 119(4):501–507.

Garg, T., Jagnani, M., and Taraz, V. (2020). Temperature and human capital in india. *Journal of the Association of Environmental and Resource Economists*, 7(6):1113–1150.

Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., and Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. *Health services research*, 49(5):1701–1720.

Giglio, L., Loboda, T., Roy, D. P., Quayle, B., and Justice, C. O. (2009). An active-fire based burned area mapping algorithm for the modis sensor. *Remote sensing of environment*, 113(2):408–420.

Giglio, L., Schroeder, W., and Justice, C. O. (2016). The collection 6 modis active fire detection algorithm and fire products. *Remote Sensing of Environment*, 178:31–41.

Glewwe, P. and Kassouf, A. L. (2012). The impact of the bolsa escola/familia conditional cash transfer program on enrollment, dropout rates and grade promotion in brazil. *Journal of development Economics*, 97(2):505–517.

Gomes-Neto, J. B. and Hanushek, E. A. (1994). Causes and consequences of grade repetition: Evidence from brazil. *Economic Development and Cultural Change*, 43(1):117–148.

Goodman, A. and Sianesi, B. (2005). Early education and children's outcomes: how long do the impacts last? *Fiscal Studies*, 26(4):513–548.

Goudarzi, G., Geravandi, S., Idani, E., Hosseini, S. A., Baneshi, M. M., Yari, A. R., Vosoughi, M., Dobaradaran, S., Shirali, S., Marzooni, M. B., et al. (2016). An evaluation of hospital admission respiratory disease attributed to sulfur dioxide ambient concentration in ahvaz from 2011 through 2013. *Environmental science and pollution research*, 23(21):22001–22007.

Graff-Zivin, J., Liu, T., Song, Y., Tang, Q., and Zhang, P. (2020). The unintended impacts of agricultural fires: Human capital in china. *Journal of Development Economics*, 147:102560.

Graff-Zivin, J., Neidell, M., and Schlenker, W. (2011). Water quality violations and avoidance behavior: Evidence from bottled water consumption. *American Economic Review*, 101(3):448–53.

Guoliang, C., Zhang, X., Sunling, G., and Zheng, F. (2008). Investigation on emission factors of particulate matter and gaseous pollutants from crop residue burning. *Journal of Environmental Sciences*, 20(1):50–55.

Halla, M. and Zweimüller, M. (2014). Parental response to early human capital shocks: evidence from the chernobyl accident.

Hansen, A. C. and Selte, H. K. (2000). Air pollution and sick-leaves. *Environmental and Resource Economics*, 16(1):31–50.

Hanushek, E. A. (1996). A more complete picture of school resource policies. *Review of Educational Research*, 66(3):397–409.

Hanushek, E. A. and Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American economic review*, 90(5):1184–1208.

Heckman, J., Ichimura, H., and Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, 65(2):261–294.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, 64(4):605–654.

Herbst, M. and Strawiński, P. (2016). Early effects of an early start: Evidence from lowering the school starting age in poland. *Journal of Policy Modeling*, 38(2):256–271.

Heyes, A., Neidell, M., and Saberian, S. (2016). The effect of air pollution on investor behavior: Evidence from the s&p 500. Technical report, National Bureau of Economic Research.

Houston, J., Allendorfer, J., Nenert, R., Goodman, A. M., and Szaflarski, J. P. (2019). White matter language pathways and language performance in healthy adults across ages. *Frontiers in neuroscience*, 13.

Hsiang, S., Oliva, P., and Walker, R. (2019). The distribution of environmental damages. *Review of Environmental Economics and Policy*, 13(1):83–103.

Huber, M., Lechner, M., and Wunsch, C. (2013). The performance of estimators based on the propensity score. *Journal of Econometrics*, 175(1):1–21.

IBGE (2013). Pesquisa nacional por amostra de domicílios, pnad. *Rio de Janeiro: Instituto Brasileiro de Geografia e Estatistica, IBGE*.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1):5–86.

INEP (2017a). Microdados da prova brasil. `http://portal.inep.gov.br/web/guest/microdados`. [Online; accessed 19-January-2019].

INEP (2017b). Microdados do enem. http://portal.inep.gov.br/web/guest/microdados. [Online; accessed 19-August-2019].

IPEA (2014). Indicadores macroeconômicos. http://www.ipeadata.gov.br/Default.aspx. [Online; accessed 15-November-2018].

Isen, A., Rossin-Slater, M., and Walker, W. R. (2014). Every breath you take–every dollar you'll make: The long-term consequences of the clean air act of 1970 (no. w19858).

Itron (2019). Blogs itron. https://blogs.itron.com/computing-a-weighted-average-wind-speed-and-wind-direction-across-multi\ple-weather-stations/. [Online; accessed 19-November-2019].

Jacob, B. A. and Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of economics and statistics*, 86(1):226–244.

Jacob, B. A. and Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, 1(3):33–58.

Jayachandran, S. (2009). Air quality and early-life mortality evidence from indonesia's wildfires. *Journal of Human resources*, 44(4):916–954.

Jimerson, S., Anderson, G. E., and Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools*, 39(4):441–457.

Jimerson, S., Egeland, B., Sroufe, L. A., and Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of school psychology*, 38(6):525–549.

Justice, C., Giglio, L., Korontzi, S., Owens, J., Morisette, J., Roy, D., Descloitres, J., Alleaume, S., Petitcolin, F., and Kaufman, Y. (2002). The modis fire products. *Remote sensing of Environment*, 83(1-2):244–262.

Kahn, M. E. and Li, P. (2019). The effect of pollution and heat on high skill public sector worker productivity in china. Technical report, National Bureau of Economic Research.

Katanoda, K., Sobue, T., Satoh, H., Tajima, K., Suzuki, T., Nakatsuka, H., Takezaki, T., Nakayama, T., Nitta, H., Tanabe, K., et al. (2011). An association between long-term exposure to ambient air pollution and mortality from lung cancer and respiratory diseases in japan. *Journal of epidemiology*, pages 1102090211–1102090211.

Kaufman, Y. J., Justice, C. O., Flynn, L. P., Kendall, J. D., Prins, E. M., Giglio, L., Ward, D. E., Menzel, W. P., and Setzer, A. W. (1998). Potential global fire monitoring from eos-modis. *Journal of Geophysical Research: Atmospheres*, 103(D24):32215–32238.

Khan, R. R. and Siddiqui, M. (2014). Review on effects of particulates; sulfur dioxide and nitrogen dioxide on human health. *Int. Res. J. Environ. Sci*, 3(4):70–73.

Khandker, S., B. Koolwal, G., and Samad, H. (2009). *Handbook on impact evaluation: quantitative methods and practices*. The World Bank.

Khaniabadi, Y. O., Polosa, R., Chuturkova, R. Z., Daryanoosh, M., Goudarzi, G., Borgini, A., Tittarelli, A., Basiri, H., Armin, H., Nourmoradi, H., et al. (2017). Human health risk assessment due to ambient pm10 and so2 by an air quality modeling technique. *Process safety and environmental protection*, 111:346–354.

Kirchhoff, V., Marinho, E., Dias, P., Pereira, E., Calheiros, R., André, R., and Volpe, C. (1991). Enhancements of co and o 3 from burnings in sugar cane fields. *Journal of Atmospheric Chemistry*, 12(1):87–102.

Koppensteiner, M. F. (2010). *Essays on human capital in Brazil*. PhD thesis, Queen Mary University of London.

Koppensteiner, M. F. (2014). Automatic grade promotion and student performance: Evidence from brazil. *Journal of Development Economics*, 107:277–290.

Lai, W., Li, Y., Tian, X., and Li, S. (2018). Agricultural fires and cognitive function: evidence from crop production cycles. *Available at SSRN 3039935*.

Lamsal, K., Jones, P. C., and Thomas, B. W. (2017). Sugarcane harvest logistics in brazil. *Transportation Science*, 51(2):771–789.

Lavy, V., Paserman, M. D., and Schlosser, A. (2011). 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):208–237.

Lechner, M. (2002). Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(1):59–82.

Leuven, E. and Sianesi, B. (2003). Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components S*, 432001.

Lin, C.-M., Li, C.-Y., Yang, G.-Y., and Mao, I.-F. (2004). Association between maternal exposure to elevated ambient sulfur dioxide during pregnancy and term low birth weight. *Environmental research*, 96(1):41–50.

Ling, S. H. and van Eeden, S. F. (2009). Particulate matter air pollution exposure: role in the development and exacerbation of chronic obstructive pulmonary disease. *International journal of chronic obstructive pulmonary disease*, 4:233.

Lipfert, F. W., Wyzga, R. E., Baty, J. D., and Miller, J. P. (2009). Air pollution and survival within the washington university–epri veterans cohort: risks based on modeled estimates of ambient levels of hazardous and criteria air pollutants. *Journal of the Air & Waste Management Association*, 59(4):473–489.

Liu, H. and Salvo, A. (2018). Severe air pollution and child absences when schools and parents respond. *Journal of Environmental Economics and Management*, 92:300–330.

Liu, Y., Paciorek, C. J., and Koutrakis, P. (2009). Estimating regional spatial and temporal variability of pm2. 5 concentrations using satellite data, meteorology, and land use information. *Environmental health perspectives*, 117(6):886–892.

Machin, S., Marie, O., and Vujić, S. (2011). The crime reducing effect of education. *The Economic Journal*, 121(552):463–484.

Machin, S. and Wilson, J. (2005). Public and private schooling initiatives in england. In *Conference on Mobilizing the Private Sector for Public Education, October*, pages 5–6.

Makles, A. (2012). Stata tip 110: How to get the optimal k-means cluster solution. *The Stata Journal*, 12(2):347–351.

Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics*, 94(2):596–606.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, 107(2):407–437.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Marques da Costa, R. M. V. (2018). *Education, Work and Crime: Evidence from Educational Reforms*. PhD thesis, University College London.

McConnell, M., Dohlman, E., and Haley, S. L. (2010). World sugar price volatility intensified by market and policy factors. Technical report.

McEwan, P. J. and Shapiro, J. S. (2008). The benefits of delayed primary school enrollment discontinuity estimates using exact birth dates. *Journal of human Resources*, 43(1):1–29.

Moretti, E. and Neidell, M. (2011). Pollution, health, and avoidance behavior evidence from the ports of los angeles. *Journal of human Resources*, 46(1):154–175.

Murnane, R. J. and Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Murnane, R. J., Willett, J. B., and Levy, F. (1995). The growing importance of cognitive skills in wage determination. Technical report, National Bureau of Economic Research.

Murphy, K. M. and Peltzman, S. (2004). School performance and the youth labor market. *Journal of labor Economics*, 22(2):299–327.

Neidell, M. (2004). Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma. *Journal of health economics*, 23(6):1209–1236.

Neidell, M. (2006). Public information and avoidance behavior: Do people respond to smog alerts? *Center for Integrating Statistical and Environmental Science Technical Report*, 24.

Neidell, M. (2009). Information, avoidance behavior, and health the effect of ozone on asthma hospitalizations. *Journal of Human resources*, 44(2):450–478.

Nilsson, J. P. (2009). The long-term effects of early childhood lead exposure: Evidence from the phase-out of leaded gasoline. *Institute for Labour Market Policy Evaluation (IFAU) Work. Pap.*

Okurut, J. M. (2018). Automatic promotion and student dropout: Evidence from uganda, using propensity score in difference in differences model. *Journal of Education and Learning*, 7(2):191.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1):152–175.

Paciorek, C. J. and Liu, Y. (2009). Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter. *Environmental health perspectives*, 117(6):904–909.

Peña, P. A. (2017). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review*, 56:152–176.

Popescu, T., Sader, E., Schaer, M., Thomas, A., Terhune, D. B., Dowker, A., Mars, R. B., and Kadosh, R. C. (2019). The brain-structural correlates of mathematical expertise. *cortex*, 114:140–150.

Provençal, S., Buchard, V., da Silva, A. M., Leduc, R., and Barrette, N. (2017). Evaluation of pm surface concentrations simulated by version 1 of nasa's merra aerosol reanalysis over europe. *Atmospheric pollution research*, 8(2):374–382.

Rangel, M. A. and Vogl, T. S. (2019). Agricultural fires and health at birth. *Review of Economics and Statistics*, 101(4):616–630.

Ransom, M. and Pope, C. (2013). Air pollution and school absenteeism: Results from a natural experiment. *Manuscript, Department of Economics, Brigham Young University.*

Reyes, J. W. (2011). Childhood lead and academic performance in massachusetts. *Childhood*, 11(3).

Ribeiro, H. (2008). Sugar cane burning in brazil: respiratory health effects. *Revista de Saúde Pública*, 42:370–376.

Rivkin, S. G. (1995). Black/white differences in schooling and employment. *Journal of Human resources*, pages 826–852.

Roderick, M. (1994). Grade retention and school dropout: Investigating the association. *American Educational Research Journal*, 31(4):729–759.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):S71–S102.

Rosa, L., Martins, M., and Carnoy, M. (2019). Achievement gains from reconfiguring early schooling: The case of brazil's primary education reform. *Economics of Education Review*, 68:1–12.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Rubin, D. B. (1980). Bias reduction using mahalanobis-metric matching. *Biometrics*, pages 293–298.

SAEB (2006). Escala saeb. ministerio da educacao. https://www.academia.qedu.org.br/prova-brasil/454-2/. [Online; accessed 15-November-2018].

Sager, L. (2019). Estimating the effect of air pollution on road safety using atmospheric temperature inversions. *Journal of Environmental Economics and Management*, 98:102250.

Sanders, N. J. (2012). What doesn't kill you makes you weaker prenatal pollution exposure and educational outcomes. *Journal of Human Resources*, 47(3):826–850.

Schlenker, W. and Walker, W. R. (2015). Airports, air pollution, and contemporaneous health. *The Review of Economic Studies*, 83(2).

Seaton, A., Godden, D., MacNee, W., and Donaldson, K. (1995). Particulate air pollution and acute health effects. *The lancet*, 345(8943):176–178.

SEB/MEC (2006). Ensino fundamental de nove anos: perguntas mais frequentes e respostas da secretaria de educação básica (seb/mec) kernel description. http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm. [Online; accessed 13-November-2018].

Setzer, A. and Sismanoglu, R. (2007). Risco de fogo: Metodologia do calculo. *Instituto Nacional de Pesquisas Espaciais*.

Sianesi, B. (2001). Implementing propensity score matching estimators with stata. In *UK Stata Users Group, VII Meeting*.

SIMA (2017). The green protocol. secretariat for infrastructure environment state of são paulo. https://www.infraestruturameioambiente.sp.gov.br/en/greener-ethanol/. [Online; accessed 10-November-2019].

Smith, J. (1996). Three essays on the economics of evaluating social programs: Dissertation summary.

Solow, R. M. (1999). Neoclassical growth theory. *Handbook of macroeconomics*, 1:637–667.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56(1):79–107.

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., Danabasoglu, G., Suzuki, T., Bamber, J. L., Bentsen, M., et al. (2018). Jra-55 based surface dataset for driving ocean–sea-ice models (jra55-do). *Ocean Modelling*, 130:79–139.

UN (2015). Millenium development goals. http://www.un.org/millenniumgoals/. [Online; accessed 18-November-2018].

UNESCO, E. f. A. P. (2007). *Education for All Global Monitoring Report 2008: Education for All by 2015. Will We Make It?* Oxford University Press.

Unicadata (2020). Observatório da cana. http://www.unicadata.com.br. [Accessed in August, 2020].

Uriarte, M., Yackulic, C. B., Cooper, T., Flynn, D., Cortes, M., Crk, T., Cullman, G., McGinty, M., and Sircely, J. (2009). Expansion of sugarcane production in são paulo, brazil: implications for fire occurrence and respiratory health. *Agriculture, ecosystems & environment*, 132(1-2):48–56.

Vanino, E. (2016). *Essays on firms innovation, internationalization and trade policy*. PhD thesis, University of Birmingham.

Webmet (2019). The meteorological resource centre. http://www.webmet.com/met_monitoring/622.html. [Online; accessed 11-November-2019].

Weinhold, B. (2008). Ozone nation: Epa standard panned by the people. https://ehp.niehs.nih.gov/doi/full/10.1289/ehp.116-a302. [Online; accessed 19-August-2019].

Wikipedia (2007). Municipalities of brazil. https://en.wikipedia.org/wiki/Municipalities_of_Brazil. [Online; accessed 15-November-2018].

Wikipedia (2019). Ozone. https://en.wikipedia.org/wiki/Ozone#Ozone_air_pollution. [Online; accessed 16-November-2019].

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wu, J., Kong, S., Wu, F., Cheng, Y., Zheng, S., Yan, Q., Zheng, H., Yang, G., Zheng, M., Liu, D., et al. (2018). Estimating the open biomass burning emissions in central and eastern china from 2003 to 2015 based on satellite observation. *Atmospheric Chemistry & Physics*, 18(16).

Zapata, J., Pont, B., Figueroa, D. T., Peterka, J., and Fraccola, S. (2015). Education policy outlook: Brazil. *OECD Publishing*.

Zhang, L., Liu, Y., and Hao, L. (2016). Contributions of open crop straw burning emissions to pm2. 5 concentrations in china. *Environmental Research Letters*, 11(1):014014.

Zhang, S., Zhong, R., and Zhang, J. (2017). School starting age and academic achievement: Evidence from china's junior high schools. *China Economic Review*, 44:343–354.

Zweig, J. S., Ham, J. C., and Avol, E. L. (2009). Air pollution and academic performance: Evidence from california schools. *Department of Economics, University of Maryland*.

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Chapter 1 - We quantify the impact of two educational policies in Brazilian public schools on (i) scores in the national examination *Prova Brasil* and (ii) the dropout and pass rates. We construct a unique longitudinal panel at the school level from 2007 to 2015, combining administrative data from several government sources. To identify the causal impacts stemming from the two policies we use Propensity Score Matching with Difference-in-differences (PSM-DID). The results reveal a positive impact of the school entrance age policy on Portuguese and Mathematics scores and on dropout rates; and a negative effect on pass rates. Although our results show a positive impact of automatic promotion on pass rates, we find no evidence that adopting automatic progression boosts the quality of education in Brazil as the dropout rates increased after that policy as well. These outcomes suggest that the two policies by themselves do not improve the quality of education in Brazil.

Chapter 2 - This chapter examines the contemporaneous causal relationship between outdoor air pollution levels and student cognitive performance in Brazil's nationwide university entrance examinations. Our analysis relies upon a unique and not yet explored student level data set, allowing us to examine the effect of particulate matter (PM_{10}) on students' scores. To this end for our main specification we construct an individual level panel data for the two days of exams across three years and apply students fixed effects to address endogeneity concerns that arise in our analysis. In addition, we take advantage of plausibly exogenous spatial and temporal variation in PM_{10} across municipalities in the states of Rio de Janeiro and São Paulo and utilise an instrumental variable approach based

on wind direction. Our results suggest that air pollution negatively impacts the cognitive performance of students with boys more affected than girls, and less well-off exam takers at the bottom of the score distribution more affected than their more privileged counterparts. *Chapter 3* - We estimate the causal effects of acute fine particulate matter ($PM_{2.5}$) fetal exposure on Brazilian students' scores in the fifth grade using the interaction between wind direction and agricultural fires as instrument for air pollution. Our reveals find that an increase of 10% of intrauterine exposure to $PM_{2.5}$ leads to a decrease of 4.6% in students' scores in Portuguese and 2.8% Mathematics. Back-of-the-envelope calculations demonstrate that a reduction of 10% of $PM_{2.5}$ has the potential to increase later life wages by 2.3%.

Acknowledgments

I am extremely grateful to my supervisors Prof. Matthew Cole and Prof. Eric Strobl for their patient guidance. Your comments were always helpful to improve the quality of this thesis. Matt, thank you so much for taking over as my first supervisor and for asking Eric to be my second. This saved my PhD. Eric, thank you so much indeed for accepting being my second supervisor in the past nine months. I have learned a lot from you in a short period of time. Thank you for the encouragement. I am in debt with both of you.

An infinite thank you to my parents and two brothers for all the support and love. You are my backbone. Phil, thank you for being my best friend in the UK and for your support and caring. Mari and Sama thank you for decades of encouragement. Thank you my two *stinkies* Benji and Lua, my best mates ever in England, my beloved dogs! I love you all dearly.

Thank you to the editor from the Journal of the Association of Environmental and Resources Economists (JAERE) Prof. Matthew Neidell and two anonymous referees who provided excellent suggestions for the paper from my second chapter.

I would like to thank some other great people who appeared in my last year of PhD journey giving me advice, encouragement and opportunities. Dr. Davide Rigo, Dr. Livia Menezes, Dr. Liyun Zhang, Dr. Martin Koppensteiner, Prof. Olmo Silva and Prof. Tatyana Deryugina.

Thanks to the Department of Economics for the scholarship and the TEDE group for course and conference financial support. Thanks Marleen for your always swift support. Thank you my office fellows Helen, Julian, Muttamas and Raeni. Although you were not

from Economics, we passed together through the highs and lows of doing a PhD. Good luck for the four of you.

Declaration

I acknowledge that a research proposal based upon some of the work (Chapter 2) within this thesis was submitted as partial fulfillment of the University of Birmingham's Advanced Research Methods module.

Contents

Acknowledgments **v**

List of Figures **xi**

List of Tables **xiv**

List of Abbreviations **xv**

1 Introduction **1**

2 Educational reforms and school quality in Brazil **5**

2.1 Introduction 5

2.2 Literature Review 10

2.2.1 Grade retention and school dropout rates 11

2.2.2 Compulsory school entrance age 14

2.3 Data Description and Policy Background 20

2.3.1 National tests - *Prova Brasil* 20

2.3.2 School Census 22

2.3.3 Descriptives 25

2.3.4 Policies background 30

2.4 Research Design 35

2.4.1 Identification Strategy 36

2.5 Empirical Findings 47

2.5.1 Compulsory school entrance age . . . 47
2.5.2 Automatic Promotion . . . 54
2.6 Sensitivity Checks and Limitations . . . 56
2.6.1 Difference-in-differences with Fixed Effects . . . 56
2.6.2 Other PSM techiques . . . 60
2.6.3 Placebo Tests . . . 60
2.6.4 Limitations . . . 62
2.7 Conclusion . . . 63
Appendix . . . 64

3 The effects of air pollution on students' cognitive performance at Brazilian university entrance tests: Evidence from changes in Wind Direction 98
3.1 Introduction . . . 98
3.2 Literature on air pollution and cognitive performance . . . 102
3.3 Data and Summary Statistics . . . 108
3.3.1 Air pollution data . . . 108
3.3.2 Weather data . . . 109
3.3.3 University entrance examination scores and other ENEM data . . . 111
3.4 Empirical Strategy . . . 113
3.4.1 Panel Fixed Effects Model . . . 115
3.4.2 The instrumental variable approach . . . 116
3.5 Empirical Results . . . 121
3.5.1 Main Results . . . 121
3.5.2 Heterogeneity across sub-populations . . . 123
3.5.3 Sensitivity Analysis . . . 129
3.6 Conclusion . . . 133
Appendix . . . 134

4 Fetal exposure to fine particulate matter and students' cognitive performance: Evidence from agricultural fires in Brazil 139

4.1 Introduction . . . 139

4.2 Data . . . 146

4.2.1 Agricultural Fires . . . 146

4.2.2 Atmospheric Conditions . . . 148

4.2.3 Air Pollution . . . 150

4.2.4 Education . . . 151

4.3 Empirical Strategy . . . 153

4.3.1 The basic OLS models . . . 153

4.3.2 The instrumental variable approach . . . 155

4.4 Results . . . 158

4.4.1 Mechanisms: Relevance of upwind fires as instrument . . . 158

4.4.2 Second Stage: Impact of air pollution on students' scores . . . 159

4.4.3 Heterogeneity . . . 160

4.4.4 School Performance and Wages . . . 163

4.5 Sensitivity Analysis . . . 164

4.6 Conclusion . . . 167

Appendix . . . 170

5 Conclusion 185

References 208

List of Figures

2.1 PSM balancing using only municipal schools (primary cycles) 46

A1 Simulated probability of completion sixth grade on time, *circa* 2005 74

A2 Brazil-Chile years of schooling per cohort of birth 83

A3 PSM balancing - state schools only - Method: Kernel - compulsory school entrance age 88

A4 PSM balancing - whole sample s=1 - Kernel - compulsory school entrance age 89

A5 PSM balancing -state schools s=1 - Kernel - compulsory school entrance age 92

A6 PSM balancing - municipal s=1 - Kernel - compulsory school entrance age 93

A7 PSM balancing - whole sample - Kernel - automatic promotion) 94

A8 PSM balancing - state schools - Kernel - automatic promotion 95

A9 PSM diagnostic balancing - whole sample s= 1 - Kernel - compulsory school entrance age 96

3.1 WSS, log(WSS), η^2, and PRE for all K cluster solutions (Optimal number of clusters 119

3.2 The effect of wind direction on air pollution 120

3.3 Residual PM_{10} and Residual Scores Scatterplot (student FE) 122

3.4 Cognitive impact of PM_{10} per distance from the station to the municipality centroid 131

B1 O_3 mean per day - IWD - 10km 136

B2 PM_{10} mean per day - IWD - 10km 137

B3 Rio and São Paulo municipalities centroids and pollution stations (raw data) 137

4.1 Heterogeneity in Particulate Matter's Impact on Students' Scores Across Portuguese Scores' Distribution 162

4.2 Heterogeneity in Particulate Matter's Impact on Students' Scores Across Maths Scores' Distribution 163

4.3 Cognitive (Portuguese scores) impact of $PM_{2.5}$ per distance from the fires points to the population's municipality centroid 165

4.4 Cognitive (Maths scores) impact of $PM_{2.5}$ per distance from the fires points to the population's municipality centroid 166

C1 Total of agricultural fires per municipality from 2000 to 2008 170

C2 Total of agricultural fires per municipality size from 2000 to 2008 170

C3 Total of agricultural fires per municipality from 2009 to 2017 171

C4 Total of agricultural fires per municipality size from 2009 to 2017 171

C5 Monthly average of counts of fires from 2001-2008 180

C6 Decomposition Additive Time Series of monthly average counts of fires from 2001-2008 181

C7 Monthly average of counts of fires from 2001-2017 182

C8 Share of sugarcane crop per municipality 183

C9 Share of rice, sugarcane, maize and wheat planted per municipality from 2000 to 2017 183

C10 Distribution of Fires, cropland and sugarcane in early 2000s 184

List of Tables

2.1 Frequency of schools taking *Prova Brasil* 22

2.2 Summary statistics for initial years per year 26

2.3 Summary statistics for final years per year 27

2.4 Descriptive statistics initial years - Policy: school entrance age 29

2.5 Descriptive statistics initial years - Policy: grade retention 30

2.6 Primary school nomenclature by the National Education Council 32

2.7 Whole sample - Treatment: compulsory school entrance age - (Kernel) . . 49

2.8 Only municipal schools - Treatment: compulsory school entrance age - (Kernel) 51

2.9 Only state schools - Treatment: compulsory school entrance age - (Kernel) 53

2.10 Whole sample - Treatment: Automatic promotion - (Kernel) 54

2.11 Only State schools - Treatment: Automatic promotion - (Kernel) 55

2.12 Only municipal schools - Treatment: Automatic promotion - (Kernel) . . . 55

2.13 DID-FE - Treatment: compulsory school entrance age (World Bank Method) 58

2.14 Treatment: Primary cycles DID-FE (World Bank Method) 59

2.15 PSM-DID estimations using placebo treatment: compulsory school entrance age 61

2.16 PSM-DID estimations using placebo treatment: automatic promotion . . . 62

A1 Marks scale for Portuguese exam in *Prova Brasil* 64

A2 Marks scale for Mathematics exam in *Prova Brasil* 65

A6 Matching diagnostic - treatment: compulsory school entrance age (Kernel method); cohort 1; s=0; only STATE SCHOOLS) 65
A3 All the PSM techniques used 71
A4 Propensity Score Matching - Difference in Differences (only states schools) 72
A5 Propensity Score Matching - Difference in Differences (only municipal schools) 73
A7 Matching diagnostic - treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; whole sample) 74
A8 Matching diagnostic - treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; only state schools) 78
A9 Matching diagnostic - Treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; only municipal schools) 83
A10 Whole Sample - Kernel - automatic promotion 90
A11 State schools - Kernel - automatic promotion 91
A12 Municipal schools - Kernel - automatic promotion 97

3.1 Summary statistics air pollution, weather and exams scores 110
3.2 The Effects of PM_{10} on Students' Scores (Main Specification with flexible weather controls) 126
3.3 Particulate Matters's Impact on Scores: Models with Instrumental Variable Wind Direction 127
3.4 Heterogeneity in Particulate Matter's Impact on Students' Scores (in SD) per gender (with bins) 127
3.5 Heterogeneity in Particulate Matter's Impact on students' scores (in SD) across sub-populations (bins) 128
3.6 The Effects of PM_{10} on Students' Scores in SD (with co-pollutant (bins)) . 132
3.7 The Effects of PM_{10} on Students' Scores in SD (Baseline Model + 1 day lead)(bins) 132
B1 Instrument for Pollution: Wind Direction – First Stage Results 135

B2 The Effects of PM_{10} on Students' Scores (Specification with flexible weather controls) . . . 136
B3 The Effects of PM_{10} on Students' Scores (Linear and no weather controls) 138
B4 Number of municipalities per distance from the monitoring stations 138

4.1 Summary Statistics for main regressions sample . . . 153
4.2 Instrument for Pollution: Count of fires per Wind Direction – First Stage Results) . . . 172
4.3 The Effects of $PM_{2.5}$ on Students' Scores - OLS and IV (Main Specification with flexible weather controls) . . . 173
4.4 Heterogeneity in Particulate Matter's Impact on Students' Scores per Gender 174
4.5 The Effects of $PM_{2.5}$ on Students' Scores (using IV with different wind directions specifications) . . . 175
4.6 First Stage Results: Instruments for Pollution using non-confidence weighted fires . . . 176
4.7 Particulate Matter's Impact on Students' Scores using fires non-confidence weighted . . . 177
4.8 The Effects of $PM_{2.5}$ on Students' Scores (SD) (without and with linear weather controls) . . . 178
4.9 The Effects of $PM_{2.5}$ on Students' Scores (SD) (with co-pollutant) 179

List of Abbreviations

API	Air Pollution Index
ATE	average treatment effect
ATT	average treatment effect on the treated
CASEN	National Socioeconomic Characterisation
CCT	Conditional Cash Transfer
CEE	State Council of Education
CETESB	*Companhia Ambiental do Estado de São Paulo*
CME	Municipal Council of Education
CNE	National Education Council
DDD	Triple-Difference in-Difference
ENEM	*Exame Nacional do Ensino Médio*
FE	Fixed Effects
GDP	Gross Domestic Product

IBGE	Brazilian Institute of Geography and Statistics
Ideb	Basic Education Development Index
Inea	*Instituto Estadual do Meio Ambiente*
Inep	National Institute of Studies and Educational Research Anísio Teixeira
IV	Instrumental Variables
IWD	Inverted Weighted Distance
MDGs	Millennium Development Goals
OECD	Organisation for Economic Co-operation and Development
PISA	Program for International Student Assessment
PNAD	National Household Sample Survey
pp	percentage points
PRE	Proportional Reduction of Error
PSM-DID	Propensity Score Matching with Difference in Differences
PTA	Parallel Trend Assumption
RDD	Regression Discontinuity Design
RE	Random Effects
Saeb	Basic Education Evaluation System
SD	standard deviation

SE	standard errors
SSA	start schooling age
UN	United Nations
UNDP	United Nations Development Program
WHO	World Health Organisation
WSS	Within Sum of Squares

Chapter 1

Introduction

This chapter introduces the research topics of the thesis, which I divide into three independent studies linked by a common core issue: human capital in Brazil. It is well-established in the literature that the economic development of a country is directly related to its level of education (Romer, 1990; Mankiw et al., 1992). Therefore, studying policies aimed at improving students' performance helps policymakers to evaluate the costs and benefits of those reforms. In addition, another relevant factor that affects cognitive performance is air pollution. Thus, this thesis casts light on these issues and assesses the impact of educational policies and bad air quality on human capital in Brazil.

In the first chapter, I investigate the effects of two Brazilian educational policies on fifth and ninth graders' scores, dropout and pass rates. In the second, I study the effects of contemporaneous effects of air pollution on the students' performance at the nationwide university entrance examination in Brazil. Finally, in the third chapter, I delve into the effects of the fetal exposure to bad air quality stemming from agricultural fires in a sugarcane crop area in the State of São Paulo on the medium and long run human capital formation.

In the first piece of research, I analyse the effect of two independent educational poli-

cies on the schools' performance at the elementary school national examination *Prova Brasil*, dropout and pass rates. The main identification challenge of this chapter is the fact that both policies were not compulsory, hence raising concerns that the results could be driven by spurious correlation stemming from self-selection. To tackle this problem, I utilise propensity score matching technique with difference-in-differences to establish the causal link between each policy on pupils' performance. To do so, I build two school-level panels using rich administrative data set from the Brazilian Ministry of Education.

As per the first policy which made it optional for schools to adopt the automatic promotion from the first to the ninth grades, I found no evidence of the impact of this policy on the Portuguese and Maths' scores at the *Prova Brasil*. Moreover, I found significant and positive coefficients for pass rates, which is the aim of the practice of non-retention. However, the results indicate also a positive relationship between non-repetition in the primary school and dropout rates, which may be signalling that the policy itself does not suffice to change the alarming dropout scenario. The second policy increased the compulsory years of schooling from 8 to 9 years, changing the starting age from 7 to 6 years old. My results indicate that this policy raised the Portuguese and Maths' scores in the national examinations, but also increased the dropout rates and lowered the pass rates.

In the second chapter, I add to a vast literature on air pollution and human capital by estimating the causal effects of contemporaneous exposure to air pollution on students' performance in university entrance exams from 2015 to 2017. To identify this relationship I use students fixed effects as main specification strategy, and, as robustness checks, I utilise a wide range of control variables, such as individual characteristics, parents' level of education, household income, weather controls (temperature, humidity and wind speed), and day and municipality fixed effects. To address possible bias in our estimations stemming from classical measurement error and omitted variable bias, besides individuals fixed effects, I also perform an instrumental variable approach where I rely on the temporal and

geographical variation in wind direction to build the instruments. The main assumption is that this variation is exogenous since it does not directly affect the students' scores, except from its effect on air pollution.

Two broad findings emerge from these results. Firstly, the effect of air pollution on exam performance appears to differ for males and females, with the performance of the former more adversely affected. This finding is consistent with previous studies including Jayachandran (2009); Sanders (2012) and Ebenstein et al. (2016). Secondly, my results suggest that the effect of air pollution on exam performance differs by socio-economic background, with poorer students scoring even worse than wealthier treated exam takers. Furthermore, I run several sensitivity checks which suggest that my results are not driven by spurious correlations and seem to add plausible evidence suggesting that cognitive performance may be hindered by bad air quality.

In the final chapter, I examine the relationship between fetal exposure to air pollution from agricultural fires and cognitive performance on fifth grade exams in Brazil. To this end I merge data from the Brazilian Ministry of Education; NASA's satellite remote sensing data with active fires detection and air pollution data and atmospheric conditions gridded data from the Japanese Reanalysis archive, JRA-55. The main identification concern of this research lies on the fact that it may be possible that unobserved time-varying features may cause endogeneity in the analysis due to omission and classical measurement error inherent to this type of air pollution study. Thus, to address this problem I rely upon changes in wind direction and fires' location as predictors of air pollution, which yield credible evidence that the exposure to poor air quality during the gestational period causes a reduction in cognitive ability of students. More precisely, I argue that students living in municipalities located upwind from the fires present lower academic performance than their downwind dwellers counterparts.

My results for this chapter coincide with the previous one in this thesis as I find evidence that fetal exposure to air pollution hinders students' performance later in life. Also, when stratifying the sample per gender and scores, I find that males and weaker students seem to be more negatively affected by poor air quality. On top of that, to check the internal validity of my findings I perform a series of tests using different sets of instruments and distances between fires and municipality's population centroid, equations with co-pollutant SO_2 and falsification tests to which our results are robust and consistently suggest the harmful effect of fetal exposure to air pollution and cognitive health. In addition, I perform back-of-the-envelope calculations showing that a reduction of 10% of $PM_{2.5}$ during the third trimester of pregnancy has the potential to increase later life wages by 2.3%.

These findings shed light on the important issue of wealth accumulation mainly in developing countries with sharper levels of inequality. Suffice it to say that the less well-off, who live in areas with low environmental standards, end up suffering more from the negative impacts of fetal exposure to air pollution on their cognitive performance, and hence deepening the wage inequality in Brazil.

The next three chapters present those three studies in detail, and the last chapter concludes this thesis.

Chapter 2

Educational reforms and school quality in Brazil

2.1 Introduction

The economic development of a country is directly related to its level of education (Romer, 1990; Mankiw et al., 1992). This is recognised in Brazil whose second Millennium Development Goals (MDGs) (UN, 2015)[1] is to achieve universal primary education. As such, school attainment, quality of schooling and school dropout deserve continuous investigation from the standpoint of public policy. According to the Brazilian Institute of Geography and Statistics (IBGE) (*Instituto Brasileiro de Geografa e Estatística*), in 2008, an adult had, on average, seven years of schooling and around 15 million Brazilians aged 10 years old or more were illiterate; despite the fact that the Law 5692/1971 had increased the compulsory schooling age from 10 to 14 years old. Due to the need of understanding the mechanisms that can lead to an upgrade in the quality of education in the country, in this chapter, I investigate the effects of two educational policies that took place in 1996 and 2006 - automatic promotion and compulsory school entrance age. I study the effects of these reforms on the national examination *Prova Brasil* scores for

[1]There are eight Millennium Development Goals proposed by the United Nations (UN), and achieving universal primary education is the second target.

Portuguese and Mathematics, dropout and pass rates.

Bruns et al. (2011) reports that education has been improved in Brazil, where the average years of education has almost doubled over the last 20 years, as has the proportion of adults who have completed secondary school. Brazil has managed to meet some of the MDGs, however its students' learning performance still lags behind wealthier nations, with many high schools losing the majority of theirs pupils before graduation, presenting historically high dropout rates. In spite of the significant net educational level of the population between 7 and 14 years old in the primary school rising from 81.2 to 97.7 percent (%) from 1990 to 2012, Brazil has a long way to enhance its education quality; in particular, grade repetition and school dropout which affect the quality of schooling and are quite high[2].

Notwithstanding high expenditures in the past two decades - data from IPEA (2014) shows that, in 2014, Brazil spent 5.6% of its Gross Domestic Product (GDP) in education; i.e., two percentage points (pp) more than the investment made in 1994 (3.8% for the latter period) -, Brazil has not showed substantial advances in the returns to education, presenting levels of productivity stagnated at the levels they were in the 1980s. In comparative terms, according to Zapata et al. (2015), the country demonstrates some improvement in the Program for Program for International Student Assessment (PISA), yet quite small in comparison with other South American countries, for instance Chile. In this particular case, Brazil currently presents the level of schooling of the 1980s of the former. The National Household Sample Survey (PNAD) (*Pesquisa Nacional por Amostra de Domicílios*) IBGE (2013) and the National Socioeconomic Characterisation (CASEN) surveys Casen (2006) indicate that Brazil took almost 25 years to reach the years of schooling of Chile, i.e., in 1988, the former country achieved the mark of 10 years

[2]As shown by the United Nations Development Program (UNDP) report from 2012, one in four students who starts elementary school in Brazil leaves school before completing the last grade, which means that Brazil has the third larger dropout rate, around 24.7% among 100 nations UN (2015).

of schooling on average, a number that the latter achieved in 1965. Only by 2013, did Brazil achieve the same average in years of schooling accumulated as Chile did in 1980[3].

In addition, the students' retention impacts children and schools as the repeater feels no longer stimulated, raising the odds of dropping out of school. When one student fails a determined grade at school, not only him and his family have to bear the costs of repetition, but so do the public schools. This happens due to the fact that the institutions have to support the repeater enrolling another time at the same grade, which diverts public resources that could be used either to improve the schools' infrastructure or to admit a new student. For instance, when investing in the failed student the school wastes inputs per student as it requires more investment in school capital, such as larger classrooms, more school meals budget, desks, computers, hiring more teachers; not to mention that having older students in the same classroom may affect the younger pupils' performance in the current grade.

Some scholars think that grade retention also negatively affects the repeaters since it stigmatises them and damage their self-esteem by being forced to be among younger classmates, which usually contribute to the pupil to dropping out from school (Jimerson et al., 2000, 2002). This situation also decreases the learning achievement and leads to the postponement of the entrance in the labour market, which is another opportunity cost that the repeater will bear. Yet, part of the literature argues that grade repetition may enhance the learning achievement by exposing students who normally have lower scores to another year in contact with the same subjects, which may boost the quality of learning and education as a whole (Gomes-Neto and Hanushek, 1994; Manacorda, 2012).

The Law 9,496 of 1996 introduced automatic grade promotion as optional in from that year, with public state schools starting to adopt it in large numbers by early 2000s as

[3]See figure A2.

one of the educational policies that the government took to meet the MDGs to extend primary education for all (universal education goal) and to lower the cost of larger pupil cohorts (UNESCO, 2007). This chapter contributes to the literature on the Economics of Education by investigating how students' performance in the national examination *Prova Brasil*, dropout and pass rates are affected by two independent policies: automatic promotion and change of school entrance age from 7 to 6 years old. My study has some similarity with Koppensteiner (2014), diverging from it, though, insofar as that author looks at the automatic promotion policy in a different period of time and just one state in Brazil, Minas Gerais. Whereas my research takes advantage of the data set feature and builds upon a panel data from 2007 to 2015 (a more recent period than the previous work), comprising the whole country at the school level. Another contribution of this work to the literature relates to the investigation of students' performance in the same classroom with the same age of entrance. It means that I am not comparing the different ages within the same grade, rather looking at similar cohorts (same entrance age) at the school level. Moreover, I add to the literature by investigating the effects of a second policy concerning school entrance age on the same response variables; a topic that has been recently analysed by Rosa et al. (2019). The authors use Triple-Difference in-Difference (DDD) as main specification strategy and find the reform to have increased 5^{th} grade students' Mathematics scores by 0.1 SD and reading scores by 0.12 SD over student gains in untreated schools. [4].

Compulsory school laws have been implemented for more than a hundred years and continue to be debated. Nonetheless, the full understanding of the implications of these laws is still pending. The literature that studies grade retention norms is far from reaching a consensus whether students, schools, society and economy will be better off should they adopt automatic promotion. Moreover, the compulsory extension of years of schooling is

[4]By the date this chapter was finalised (December 2018), I was not aware of any work investigating the school entrance age policy in Brazil. It turns out that a paper was published in February 2019 analysing exactly the same topic (Rosa et al., 2019).

usually beneficial for the whole society. According to the law 11,274 of 2006, the 6 year old students should be registered at the new first grade, while the older ones should be enrolled in the former first grade[5]. Thus, in principle, older and younger children will not be in the same classroom in the first grade; which eliminates the possibility of comparison between old and young students in the same class. Therefore, my study focuses on the impact that the younger cohorts will have on the overall school performance at the school level, especially because there is no data available on the students' marks individually. The scores are released at the school, municipality and state level.

The extension of primary education to nine years of duration, with the compulsory enrolment from the six years of age, has been a goal pursued by the national education policy for many years. The nine-year primary school has extended the time of the initial years from four to five years (the final year phase consists of 4 years), to give the child a longer period for self-learning, including literacy. Besides this, the measure was taken to decrease the gap between the students from private and public schools since some private institutions had already adopted the 6 years old rule even before the law of 2006. Understanding how those two educational policies will impact the performance, the truancy, the dropout and the pass rates is the main contribution of this chapter. I quantify the benefits that dropouts forego by leaving school before the compulsory age.

To measure the impacts of both reforms on the tests scores, dropout and pass rate, I build a panel data from 2007 to 2015 at the school level, apply Propensity Score Matching with Difference in Differences (PSM-DID) technique and find evidence that automatic promotion by itself is not able to diminish dropout rates in Brazil. I find no clear evidence that non-retention is capable of tackling the quality of education in Brazil. Regarding the second policy - compulsory school entrance age - , I show evidence that Portuguese and

[5]Previously, the students were enrolled in the first grade with seven years old. I mention "new first grade" to refer to the grade to which students enrol with six years old. This distinction is important since it is not only a matter of nomenclature, but also different educational subjects taught in the new first grade.

Maths scores improved for the cohorts that were affected by the reform. For example, for the cohorts investigated in this research, my results suggest an increase in Portuguese scores between 0.35 and 0.41 SD; and between 0.14 and 0.29 of a SD in Mathematics. Notwithstanding, my findings indicate an increase in dropout rates of between 1.46 and 2.61 pp, and a fall in pass rates. These results may indicate either that the younger children are less mature and less prepared to face the first grade at the age of 6, or that the schools and staff are not well equipped to receive the younger pupils at a more demanding grade, which made more difficult to students to pass, which can be a good sign that the policy is indeed transforming, even slightly, the quality of education in public schools. Furthermore, regarding the decrease in pass rates, the *Prova Brasil* is probably easier than the internal schools Portuguese and Maths tests, leading children to fail internal exams, even though they present better marks in the two subjects in the national examination.

The rest of this chapter is structured as follows. In Section 2.2 I present the literature review. In Section 2.3 I describe the data used in this study. In Section 2.4 I present the research methodology used to assess the causal effects of the two laws on the students performance at the national examination *Prova Brasil*, dropout and pass rates. In Section 2.5 I report the main results of this empirical work. In Section 2.6 I provide sensitivity analysis. In Section 2.7 I discuss limitations and conclusions of this study.

2.2 Literature Review

The human capital approach developed by Becker (1962) is the theoretical backbone of the Economics of Education. Evaluating the returns of investment in one more year of schooling and/or the cost of grade retention in public schools to the government generates a lively literature on this economics branch. The analyses of compulsory schooling

entrance age and grade retention both theoretically and empirically casts light on the impact of these two policies on students performance, dropout and pass rates. Some other works look at the effects of these reforms on the labour market and on crime rates. Most of these studies agree that increasing the period that the children spend at school boosts their academic scores, their productivity, generating the possibility of higher salaries in the long run; or even lowering the crime rate via the incapacitating effect from taking children out of the streets as Machin et al. (2011) demonstrate.

In the next sections, I review the existing literature on compulsory schooling age and grade retention, assessing their effects on the schools performance on the national examination *Prova Brasil*, dropout and pass rates.

2.2.1 Grade retention and school dropout rates

There is no consensus in the literature on grade retention and its effects on students performance, dropout and pass rates. Some studies show that non-repetition policies lead students to be more self-confident and less stigmatised for being older than the average age within a classroom or a cohort; hence, they are less likely to dropout school (Roderick, 1994). On the other end of the spectrum, there is research showing that grade repetition enhances the quality of education, which is, in the first place, the core objective of any school reform (Gomes-Neto and Hanushek, 1994). The literature shows mixed results with sometimes positive and negative relationship between grade retention and school dropout, pass rates and exams performance. This conundrum is usually observed in developing countries, which present high retention rates and low investment in public education. Last but not least, there are also some works that point out the negative peer effect stemming from the older students (the repeaters) towards the younger (see for example Manski (1993); Lavy et al. (2011)).

Grade retention - the situation in which a student has to repeat a school year for having failed it in the previous year - from kindergarten to sixth grade is associated with a substantial increase in the probability of dropping out even after controlling for differences in background and post-repetition grades and attendance (Koppensteiner, 2014). For the United States, Roderick (1994) found that pupils who were overage at sixth grade were more likely to drop out of secondary school[6]. The author estimated that approximately one quarter of students left school, and those who remained had significant declines in attendance. These findings corroborate the thesis underpinning the present paper that automatic progression - that is, allowing students to progress from one grade to the next regardless of their academic performance during the previous year - can reduce the dropout rates ; i.e., percentage of students who leave school before the school year ends, leading, eventually, to a long run increase in the years of schooling.

Still, in a work on Chicago public schools, Jacob and Lefgren (2009) used a Regression Discontinuity Design (RDD) to assess the impact of grade retention on high school completion. The authors observed that grade failure leads to a modest increment in the odds of dropping out for older students (eight graders), whilst, for younger pupils (sixth graders), no significant effect was detected. The authors also examined the impacts of grade retention separately on those students who ultimately drop out and found no evidence that repetition changes the timing of dropout. Nonetheless, the problem of selection bias in the sample was underscored, which may have caused these counter-intuitive results.

The impact of automatic promotion on students dropping out of Uganda's primary education was investigated by Okurut (2018) using propensity score and difference in differences technique. The author found evidence that the practice managed to decrease dropout rates for those children. In examining Brazil's southern neighbour Uruguay, Manacorda (2012) worked with administrative longitudinal micro-data on secondary pupils

[6]In Brazil, secondary school is synonym of high school.

between 1996 and 1997 and, using a fuzzy RDD, the author observed that grade repetition leads to substantial dropout rate and lower educational attainment even four to five years after the grade failure first occurred. This piece of evidence builds the assumption I will make to create a dichotomous treatment variable - repetition or automatic progression - so as to evaluate whether the dropout rate will ultimately shorten due to the law of non-repetition in the early years of primary school.

Gomes-Neto and Hanushek (1994) conclude that Maths performance appears to improve more than Portuguese through grade repetition in Northeast Brazilian public schools; however, delayed promotion is quite costly - considering a full cost of 2 years of schooling. As such, the second best policy would be the mandatory promotion for the country's public school system. Whereas, the first best policy suggested is to enhance the quality of primary schools to increase the students achievement directly. Manacorda (2012) also argues that, regardless of the costs of grade retention, failing students may prevent the quality of education to decrease due to the deterrence effect of grade retention, which plays the part of an intimidating penalty that ends up inciting students to study harder to avoid failing the grade. Therefore, the threat of repetition produces positive effects not only to potential repeaters, but also to other low performance students (Koppensteiner, 2010). Consequently, new hybrid policies have appeared in the educational scheme to circumvent such a trade-off. For instance, policies that blend grade failure with remedial education, for instance summer schools, individual tutoring, support classes, etc (for more on remedial education, see (Jacob and Lefgren, 2004)). As shown in Banerjee et al. (2007) with respect to India, there were considerable pay-offs from the implementation of informal and cheap remedial education among less well-off children in that country.

All in all, this can be a plausible alternative to the grade retention - which can be an expensive policy - for other developing countries, especially for Latin American nations, whose repetition rates are higher than the Organisation for Economic Co-operation and

Development (OECD) countries, and have been repeatedly appearing at the bottom of the PISA ranking. In Brazil, there are similar policies to tackle low quality of education going on recently; for example, compensation policy - Conditional Cash Transfer (CCT)[7] such as *Bolsa Escola/Bolsa Familia* (¡for the impacts of these Brazilian CCT on education, see) Glewwe and Kassouf (2012); and *Progresa/Oportunidades* in Mexico; the P-900 in Chile, among others. *Bolsa Escola/Bolsa Familia* has shown a positive impact on the school attendance and negatively the dropout rates of the children whose families received the money.

Concerned with the quality of education after the law of non-failure, Ferrão et al. (2002) investigated the impact of that policy on students' performance from two states in Brazil: Minas Gerais and São Paulo. Using multilevel regressions, the authors found that there is no substantial loss of quality in education due to the implementation of non-repetition policies. Covering Brazil, Koppensteiner (2014) examines the effect of automatic grade progression on academic achievement in 1,993 public primary schools in the state of Minas Gerais, in Brazil, using a propensity score matching (PSM) and difference-in-differences (DID) approach. The author found a negative and significant effect of about 7 percent of a standard deviation on Math test scores; evidence in support of the disincentive effect of automatic progression. That study also found non-significant effects of the automatic progression to the dropout rates 2006 to create his treatment and control groups. Notwithstanding, the author just focused on Minas Gerais, one of the wealthiest and most educated states in the country in disregard of other regions' idiosyncrasies.

2.2.2 Compulsory school entrance age

Rosa et al. (2019)'s recent work use DDD as main specification strategy and find the reform to have increased 5^{th} grade students' Mathematics scores by 0.10 SD and read-

[7]CCT consists in giving money to the poor, conditionally to following pre-defined behaviour set by the government.

ing scores by 0.12 SD over student gains in untreated schools. Also, the literature on the topic has looked at other countries' schooling expansions using mostly student level data. Very little has been done related to the increase in years of schooling by lowering the entrance age as it happens in Brazil. For instance, in Herbst and Strawiński (2016), they use PSM-DID to investigate the reduction from 7 to 6 years old to start school in Poland, finding that older students outperform their six-year-old counterparts in writing and Maths by 0.11 ad 0.04 standard deviations, and the same performance for reading is observed.

Oreopoulos (2006) uses DID-FE techniques to assess the overall impact of compulsory schooling laws in Canada between 1920 and 1990. The author sheds light on the historical changes to compulsory schooling laws in that country and its effects on the would-be-dropouts under different circumstances than those that took place in other countries. Firstly, his main findings show that mandating education substantially increased adult income and substantially decreased the likelihood of being below the low-income cut-off unemployed, and in a manual occupation. These outcomes indicate significant gains from this sort of legislation, which seem unlikely to be offset by the costs incurred to remain in school. Secondly, the author uses changes in compulsory school laws to estimate the lifetime opportunity costs associated with dropping out of high school. Likewise, if the returns to schooling are very high, this might imply that the financial constraints or psychological costs associated with going to school must also be very substantial among dropouts. Conversely, if these explanations build upon constraints and costs that appear unlikely to hold, then they do not suffice to explain the dropout behaviour. That is, there might be other factor(s) that may be causing the truancy and the school evasion. For example, teenagers usually make their decisions based on cultural or peer influences, not to mention that youngsters tend to undervalue the returns from staying in school or even to ignore the future consequences of their decisions (Akerlof and Kranton, 2002). Therefore, these reasons and the empirical outcomes of Oreopoulos (2006) suggest that

compulsory schooling laws in Canada diminished the dropout rates.

SSA laws make parents and educators face an important trade-off. Usually, although those pupils who are the youngest in their school cohort have, on average, lower learning performance, they also have slightly higher educational attainment. Documenting evidence from a large cohort in California and Texas, Dobkin and Ferreira (2010) used RDD and observed that the school entry laws not only increased educational attainment of students who enter school early, but also lowered their academic performance while in school. Despite these results, they found no evidence that the age at which children enter school influences job market outcomes, such as wages or the likelihood of being employed, suggesting that the net effect on adult labour market aftermaths of the increased education attainment and poorer academic performance is close to zero.

Still delving into the North American continent educational laws, Peña (2017) thoroughly examined an unanticipated policy reform that shifted by four months the cut-off date for school eligibility in the state of Tlaxcal, Mexico. The author tested the effects of relative age on exam scores in grades 3 to 9, and observed that different identification strategies produce similar results: 1 year of additional age confers an advantage of roughly 0.3 standard deviations. First, using 2009 and 2012 data the author compared relative age and test scores between cohorts that started school right before and right after the reform with using DID across seasons of birth. Second, he used RDD using as the switch point the new unanticipated cutoff date for school enrolment on the 2012 data. Finally, the author instrumented relative age with assigned relative age using the 2012 data. By contrasting Tlaxcala with a neighboring state that did not change the cutoff at the same time, the study decomposed the effect of relative age into a positive effect of age at test, and a negative effect of the position in the distribution of age (for more on age at test and age at school entrance impacts on learning and attainment performance, see below) (Crawford et al., 2007, 2014). The research shows that older pupils have posi-

tive spillovers, whilst younger classmates present negative spillovers. As such, were both groups tested at the same age, the youngest would outperform their older counterparts.

Regarding South America, Berlinski et al. (2008) shed light on the effect of pre-primary education rapid expansion on pupil's ensuing school outcomes in Uruguay. The authors used a within household estimator and found small gains from pre-school attendance at early ages, which were boosted as children grew up. The work also shows that, by 15 years old, treated children have accumulated 0.8 extra years of education and these students were 27 percentage points more likely to be still on the school benches than their untreated siblings. In addition, the authors used Instrumental Variables (IV) estimates to control for self-selection of siblings into pre-school and found analogous results. Therefore, pre-primary schooling seems to be a successful and cost-effective policy to prevent early grade failure, suggesting long lasting repercussions in low income countries as it is the case of Uruguay. On the other end of the spectrum, in Chile, a one year delay appears to decrease the probability of repeating first grade and increases fourth and eighth grades test scores, according to McEwan and Shapiro (2008). The authors looked at administrative data on Chilean students that includes exact birth dates, allowing them to apply RDD to estimate the effect of delayed school enrollment on pupil results. Based on enrollment cut-offs, the study shows that a one-year postponement in school admission lowers the probability of failing the first grade by two percentage points; it also increases fourth and eighth grade test scores by more than 0.3 standard deviations, with larger effects for boys.

Goodman and Sianesi (2005) investigated the effects of taking early education - before the compulsory starting age of 5 - and of pre-school on a cohort of British children born in 1958[8]. Controlling for child, parental, family and neighbourhood characteristics, they found some positive and long-lasting effects from early education. Particularly, pre-compulsory education appears to significantly upgrade cognitive tests at age 7 at a

[8]Excluded Scotland.

small magnitude, remaining significant up to age 16. Whereas, attendance of pre-school - nursery or playgroup yields a positive though short-lived impact on test scores. Not only better scores are positive outcomes from compulsory schooling age reform, but also socialisation effects are found positive and short-lasting.

An important aspect of start schooling age (SSA) policies is covered by Crawford et al. (2007), who argue that it is the age at the tests effect that matters the most, instead of the SSA. Implementing RDD, the authors examine the impact of month of birth on national achievement test scores, finding small negative effect of SSA at age of 7, but this impact is dwarfed by the age-at-exam effect. Furthermore, they found that the SSA or length-of-schooling effect has disappeared by age 14. The authors complemented that work with another study Crawford et al. (2014) and using RDD again they estimated an upper bound of the effect of age at test. They found that the difference in age at which cognitive skills are tested accounts for the majority of the difference in the outcomes above.

Looking at Norway's SSA reform, Black et al. (2011) found small and negative effect of starting school older, although much larger positive effects of age at test. Besides that, starting older leads to lower earning until the age 30. The authors also found little impact of SSA on education attainment, but boys who start older are less likely to have poor mental health at the age 18; as well as negative effect of starting older and teenage pregnancy. Their analysis was build up at the students level, with a very defined cut-off date and change in trajectory impacts, allowing the use of a RDD to delve into the aftermaths of that SSA reform. Notwithstanding these findings, another study on Norwegian compulsory kindergarten effects stemming from an SSA changing the age from 7 to 6 year old did not find any effects on later school performance (Drange et al., 2012).

Studying Germany SSA, Fertig and Kluve (2005) used PSM-DID to capture the effect of age at school entry on educational attainment. The author found a negative rela-

tionship between the age the children start school and their learning outcomes, both in terms of school degree and probability of grade repetition and higher likelihood of dropout school. More recently, Betthäuser (2017) assessed the effects of compulsory schooling law change from 8 to 9 years in Germany by using DID and found evidence that the reform narrowed the gap in educational performance between different social origin groups. As a consequence, the reform also produced impacts in the long run as the author observed a reduction in the inequality in the labour market, leading to a higher social mobility in the country.

Interested in the effects China Compulsory Education Laws from 1986, Zhang et al. (2017) estimate the effect of school starting age on academic achievement on junior high school students. The authors apply RDD to estimate the treatment effects on enrolling at school at a cut-off point, by comparing the outcomes near the threshold, which is the child's age on September 1st in the school year that she entered primary school. The authors demonstrated that one year delay is associated with a decrease in with a 0.303 standard deviations of cognitive scores, oppositely to many Western countries results. However, this negative effect is caused by human capital accumulation prior to primary school entry. Firstly, the authors demonstrated a larger negative effect for children from rural schools and a small effect for those from urban schools. Secondly, the SSA is negatively related with the length of pre-school attendance. The author conclusions suggest that wealthier families may be able to better offset any negative impacts from starting school at a younger age by investing more in children before they enter primary school. Moreover, due to the lack of pre-schools in China, poor families are more prone to keep their children at home before they can start primary school. While, the children from richer households live in a more stabilised family environment and attend higher-quality pre-school. Hence, along with the preference for early school enrolment, the entrance cut-off aggravates the socio-economic gap in school performance.

2.3 Data Description and Policy Background

For this chapter, I build a biennial panel using administrative data at the school level from 2007 to 2015 with 55,567 primary public schools located in Brazil to assess how the implementation of two educational policies affected the test scores of municipal and state schools in the *Prova Brasil* examinations; the pass and the dropout rates. I merge different data sets from various official websites to collect the dependent, the main explanatory variable - the treatment - and the other control variables as described in the next subsections.

2.3.1 National tests - *Prova Brasil*

The scope of the *Prova Brasil* is to evaluate the quality of education in public schools in Brazil, comprising Portuguese and Mathematics tests and a socioeconomic questionnaire, taking place every two years (in the odd years). To minimise costs and make the evaluation easier for the Ministry of Education, not all the public schools in the country take part in this examination. Besides this, the tests are not compulsory, i.e, there are schools and students that do not participate. to take part in the tests, schools must have at least 20 students enrolled in the assessed grades. The pupils should pass a Portuguese and a Mathematics test, at the 5^{th} and 9^{th} grades, and the scores are given at the school level, per states and per municipalities.[9]

The *Prova Brasil* is created by teachers and researchers from the whole country, who elaborate questions following specific instructions from the Ministry of Education. These guidelines present the set of skills that should be assessed by the tests per grade and disciplines. Unlike the exams applied by the teachers in the schools, the methodology adopted by the *Prova Brasil* addresses the evaluation of the schools' and the educational system's

[9]Brazil is a federation composed of the union of the 26 states, the Federal District, and the 5,570 municipalities. The latter are administrative divisions of the Brazilian states which have, on average, 214 municipalities (Wikipedia, 2007).

quality rather than the students individually. Likewise, the guidelines and the Basic Education Evaluation System (Saeb)[10] marking scale (see Tables A1 and A2) are the same for all the editions of the exams allowing the comparison between results from different years.

Overall, there are 21 different types of questions booklets. Considering that the classes have approximately 21 students on average, pupils end up not answering the same questions, which lessens the chances of students be prepared by their teachers for the tests. The questions are of multiple choice, with 4 or 5 items each, and only one correct. The scores are divided by skills levels; ranging from 0 to 425 marks (elementary school goes up to 400) for Portuguese, and from 125 to 475 (elementary school up to 425) for Mathematics. Table 2.2 summarises the records of treated and untreated schools in the *Prova Brasil* Mathematics and Portuguese tests.

From 2007 to 2015, 55,567 public schools participated of the *Prova Brasil* for fifth graders per year; while only 40,334 schools took the tests for the ninth year. This is due to the fact that there are fewer schools for the last years of primary school than for the first years. Moreover, considering the increase in dropout rates after the 5^{th} grade, fewer students take part in the national examination at the ninth grade. Still, the same schools that took the exam for the the 5^{th} grade do not necessarily take it for the the 9^{th} grade in the same year. Hence, if I merge the data set with the *Prova Brasil* results from initial years (the 5^{th} grade) with the final years (9^{th} grade) in the same panel, I end up with only 26,511 schools that took both exams in the same year as demonstrated in the Table 2.1 below.

In light of that, from the *Prova Brasil* data set, I take the Portuguese and Maths results and pass rates to use as dependent variables in this study. Thus, to keep as much

[10]Saeb is composed of a set of large-scale external evaluations that allow the National Institute of Studies and Educational Research Anísio Teixeira (Inep) (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*) to make a diagnosis of Brazilian basic education and some factors that may interfere with the student's performance, providing an indication of the quality of the education offered.

Table 2.1: Frequency of schools taking *Prova Brasil*

	5th grade	9th grade	both exams
Year		Frequency	
2007	55,567	40,334	26,511
2009	55,567	40,334	26,511
2011	55,567	40,334	26,511
2013	55,567	40,334	26,511
2015	55,567	40,334	26,511

Note: Table elaborated with information from Basic Education Development Index (Ideb) (*Indice de Desenvolvimento da Educacao Basica*) microdata. Note that there are more schools registered to take the fifth year exams than the fifth year, and that the number of schools that took both exams in the same year drops by around half.

data as possible, I create three different panels: (i) one with just initial years (4^{th} or 5^{th} grades), which I merge with the data set from school census - which bring information on schools, teachers and students - just for first years; (ii) a second one with the same information, but for final years (8^{th} or 9^{th} grades); and (iii) a third panel with less data as it includes only the schools that took the *Prova Brasil* at initial and final years. I also use the dropout rates as response variable, which I collected from the school census and merged at the school level with these panels.

2.3.2 School Census

The Inep, an agency linked to the Ministry of Education, carries out the School Census, an annual nationwide statistical-educational survey. The census inquires into basic education information, such as number of enrolments; pass and dropout rates; students income; pupils' parents' (or legal tutors') characteristics; schools' infrastructure and teaching functions. The data are provided by the public and private schools themselves, state and municipal education systems. This unique and rich data set provides the dropout rates - herein used as another dependent variable - as well as the control variables adopted to conduct the propensity score matching and the econometric models. For example, par-

ents' educational level, gender, age in the date of the examinations, children work and household details.

Treatments: Compulsory school entrance age and automatic promotion

The school census provides the data for the two treatment variables, each of them accounting for the automatic promotion and for the change in the compulsory school entrance age. The survey asks the schools if they adopt the non-retention policy in the year it is being conducted; and if the school opted to enrol children aged 6 years old in the first grade. The non-repetition policy is optional; that is, the educational systems municipal or state decide whether the schools under their control will opt for automatic promotion, and the duration of this decision. As a result of the decision being made by the system rather than by the schools themselves, the concern regarding the possibility of endogeneity stemming from that option can be relaxed here.

As per the change in the compulsory school starting age, the government allowed a period of transition and adaptation for the schools to welcome the younger pupils. Thus, from 2006 to 2010, the state and municipal schools could choose if they would receive the younger students in the first grade; and, from 2010 onwards, the new starting age of six years old started to be compulsory for all school in the country. The optional phase is of the utmost importance for the public schools since it enabled them to train the teachers to cope with younger pupils in the first grade; to adapt the classrooms, furniture and course materials suitable for the younger students cohorts.

Variables for matching and control

I use a number of variables to carry out the propensity score matching and to control schools' idiosyncrasies for as many observables as possible and to try and minimise the

possibility of omitted variable bias. The following variables were also included in the models to identify the common pre-trends trajectory, which is a crucial assumption of the difference-in-difference technique: students' characteristics per school: gender; age; pre-school; race; work; ever failed; live with parents; mother schooling; parents' interest in children education (see Appendix for more detailed variables description).

To transform the individual-level variables into proportions per school, I utilise student level data, by collecting the data from the socio-economic questionnaires from *Prova Brasil*; and from the School Census students data set. I collapsed by proportion each item of the questionnaires and transformed them into a pupil level panel data so as to merge with the school level panel data. Hence, I created the proportions of each characteristic of interest per school, accordingly. Despite the existence of such information at the student level, there is no information at the same level for the marks, pass and dropout rates, which is the reason why I could not merge and analyse this entirely at the student level.

After merging the data sets, I remove from the sample all the inconsistent observations and missing or incomplete data. Furthermore, I drop the federal schools - schools that receive funding from the federal government, hence wealthier than state and municipal funded schools - from the samples as they traditionally display higher scores than the other public schools and would differ from the rest of the sample. I also dropped the high schools, the special education, the distance learning, and the adults education so that the remainder observations are only the state and municipal schools, with presential 5^{th} and 9^{th} grades. Private schools are not included in my investigation because they were not obliged to pass the national examinations - *Prova Brasil*.

2.3.3 Descriptives

The *Prova Brasil* and the Ideb data sets bring the examination results separating the schools into two groups: initial years, from first to fifth grades; and final year, from sixth to ninth grades. Following this structure, I split the main summary statistics into Tables 2.2 and 2.3. Table 2.2 demonstrates that the pass rates increased over the period from 2007 to 2013, though, from 2013 to 2015, there was a drop of approximately 13.71 pp. The same pattern is seen for Portuguese and Maths scores: the marks increased in the same interval of time and decreased from 2013 to 2015 (Portuguese around 3.89 pp lower; and Maths, 5.6 pp). Regarding the dropout rates, similar movement occurred: there was a negative trend between 2007 and 2013, with a change of slope to an increase of the dropout rates from 2013 to 2015 of roughly 10.69 pp.

Note that the number of schools change over time and across response variables. This is due to the fact that not all the same schools take exams every year and not all of them have their scores revealed. Besides this, the dropout and pass rates come from a data set different from the *Prova Brasil*'s, which means that some observations were dropped when merged as sometimes the schools have their scores available for one year in one data set but do not have for pass and dropout rates. In addition, the number of schools that take the examinations in the fifth grade is larger than the ninth grade. All the data sets I used to match show fewer ninth grade schools than fifth grade ones, certainly because there are fewer schools for the higher grade as there are fewer students that achieve the ninth year (mainly because the dropout rates increase after the fifth grade). Another explanation, for example, is that one school may have taken the fifth grade exam in a given year, though not the final exam in that same year.

When it comes to final years, pass rates have increased across the whole period and dropout rates have dropped from 2007-2013, but slightly increased by 7.23 pp from 2013-2015. The performance in Maths and Portuguese have oscillated throughout the years as

Table 2.2: Summary statistics for initial years per year

Initial yrs		2007	2009	2011	2013	2015
Pass	Mean	86.37	87.58	90.17	91.55	90.11
	Sd	10.67	10.06	8.66	7.84	8.21
	Min	0.9	0	0	29.6	19
	Max	100	100	100	100	100
	Obs	35,855	42,109	39,198	37,601	22,698
Dropout	Mean	2.95	2.35	1.67	1.31	1.45
	Sd	4.88	3.99	3.03	2.38	2.44
	Min	0	0	0	0	0
	Max	81.8	100	100	31.8	39
	Obs	35,905	42,187	39,447	30,839	22,698
Port	Mean	169.73	192.33	193.71	197.94	190.25
	Sd	18.77	21.5	22.13	25.96	23.85
	Min	103.67	121.64	106.06	104.95	106.06
	Max	261.57	243.33	243.98	315.74	314.75
	Obs	28,650	34,574	32,622	30,839	17,311
Maths	Mean	187.17	212.56	211.8	214.58	202.54
	Sd	20.56	28.15	29.31	35.8	23.17
	Min	118.17	104.97	104.97	10.59	133.86
	Max	320.88	342.32	339.22	342.32	342.32
	Obs	28,650	34,572	32,622	30,847	17,311

Note: This Table shows contains the summary statistics for the dependent variables Pass and Dropout rates, Portuguese and Maths' scores. Initial years comprise from first to fifth grades.

shown in Table 2.3.

Table 2.3: Summary statistics for final years per year

Final yrs		2007	2009	2011	2013	2015
Pass	mean	78.49	79.83	82.34	83.46	83.86
	sd	12.57	12.23	11.99	11.35	11.37
	min	12.7	0	0	0	0
	max	100	100	100	49.3	100
	obs	15,561	17,168	16,040	14,415	22,854
Dropout	mean	6.57	5.4	4.38	3.46	3.71
	sd	8.02	6.67	5.56	4.25	6.1
	min	0	0	0	0	0
	max	70.1	100	100	49.3	100
	obs	15,617	17,213	16.098	14,415	11,685
Port	mean	225.09	243.4	236.95	236.1	235.79
	sd	18.44	7.65	15.69	14.87	15.93
	min	139.6	232.2	157.4	156.4	156.4
	max	313.9	249.2	279.9	281.1	281.7
	obs	15,561	17,166	15,954	14,416	20,974
Maths	mean	237.16	247.96	251.94	247.36	247.3
	sd	18.44	8.82	19.08	15.8	17.46
	min	139.6	239.84	180.04	189.9	189.9
	max	313.9	260	318.3	309	309.5
	obs	15,561	17,166	15,954	14,416	20,974

Note: This Table shows contains the summary statistics for the dependent variables Pass and Dropout rates, Portuguese and Maths' scores. Final years comprise from sixth to ninth grades.

Additionally, Table 2.4 brings the behaviour of the response variables separated by treatment and control groups (policy: school entrance age) and per year between 2007 and 2015. The policy is from 2006, but there is no data available on the response variables for that year; therefore, I start my panel from 2007. I consider the first cohort to be affected by the new school entrance age the one which starts school at the age 6 in 2007 and take the fifth grade national exams in 2011. Hence, the effects of the policy can be noticed in 2011 for the first cohort doing the fifth year tests, and in 2015 for the same cohort doing the exams at the ninth grade. Consequently, one cannot observe any impact of that policy before the year 2011 as all students that took the exams in 2007 and 2009 are not affected by the lower entrance age, i.e., they started school at the age of 7. Also,

Table 2.5 brings the summary statistics for the treated and control groups with respect to the grade retention policy; where, retention refers to the control group and non-retention, for the treated schools. The figures and tables indicate a change in the means after the implementation of the policy.

Table 2.4: Descriptive statistics initial years - Policy: school entrance age

Year		2007		2009		2011		2013		2015	
Dep.Var.	Yrs.sch	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
	8 years	84.54	12.49	82.99	12.74	83.56	12.96	83.9	11.37	86.87	9.69
Pass	Obs	12,197		5,920		632		89		403	
	9 years	87.32	9.46	88.33	9.34	90.27	8.52	91.57	7.82	90.16	8.17
	Obs	23,658		36,189		38,566		37,512		22,295	
	8 years	3.46	5.72	3.89	5.45	3.97	7.07	2.92	4.3	2.44	3.54
Dropout	Obs	12,217		5,937		644		89		403	
	9 years	2.69	4.35	2.1	3.63	1.63	2.91	1.31	2.37	1.43	2.41
	Obs	23,688		36,250		38,803		37,512		22,295	
	8 years	168.84	18.3	184.18	19.56	182.93	23.32	175.65	18.71	180.36	23.67
Port	Obs	11,325		5,793		622		88		386	
	9 years	170.31	19.05	193.97	21.49	193.92	22.05	198	25.95	190.48	23.81
	Obs	17,325		28,781		32,000		30,751		16,925	
	8 years	186.2	20.14	204.05	22.89	197.46	28.19	197.4	19.72	193.77	22.26
Maths	Obs	11,325		5,792		622		88		386	
	9 years	187,81	19.05	214.28	28.79	212.08	29.27	214.63	35.82	202.74	23.15
	Obs	17,325		28,780		32,000		30,759		16,925	

Notes: Pass and dropout rates range from 0 to 100 in percentage. Port stands for Portuguese results and range from 0 to 400 marks. Maths results range from 125 to 425 marks. Yrs.sch stands for the duration of schooling: 8 years non-treated and 9 years treated.

Table 2.5: Descriptive statistics initial years - Policy: grade retention

	2009		2011		2013		2015	
Treat/control	mean	sd	mean	sd	mean	sd	mean	sd
retention	85.1	10.26	88.23	8.87	90.08	7.9	89.84	8.13
Obs	26,166		23,974		22,995		16,254	
non-retention	91.64	8.24	93.21	7.35	93.86	7.15	90.76	80.36
Obs	15,943		15,224		14,606		6,444	
retention	2.89	4.38	2.05	3.37	1.56	2.56	1.47	2.48
Obs	26,226		24,180		22,995		16,254	
non-retention	1.46	10.26	1.08	2.3	0.92	2	1.38	2.35
Obs	15,961		15,267		14,606		6,444	
retention	185.16	20.35	188.02	21.91	191.5	25.23	188.17	23.83
Obs	19,574			18,287	17,354		11,620	
non-retention	201.7	19.23	200.98	20.2	206.23	24.5	194.5	23.33
Obs	15,000		14,335		13,485		5,691	
retention	204.37	24.39	204.65	27.64	209.88	10.59	200.43	23.06
Obs	19,575		18,829		17,355		11,620	
non-retention	223.25	29.12	220.92	28.85	220.63	39.98	206.85	22.81
Obs	14,997		14,333		13,492		5,691	

Notes: Retention stands for the schools that did not opt for automatic promotion; and non-retention for school that did adopt it.

2.3.4 Policies background

The law 11,274/2006 has defined an optional period from 2006 until 2010, when schools could opt to follow it or not. The optional period was set to allow the schools and the public educational system to meet all new entrants' demand, adequacy of physical and pedagogical material, number of teachers and administrative support, appropriate training and career staff plan. Most public schools need to adapt not only their infrastructure, but also their staff to receive the younger pupils in a grade that used to welcome seven-year-old children. This includes the re-organisation of elementary school, for first and nine years.

Compulsory school entrance age

The law 11,274/2006 sets out an optional period from 2006 until 2010, when schools could decide whether they would enrol six-year-old children in the 1^{st} grade. The new cut-off date for children to start the primary school is 6 years old completed or to be completed by the start of the school year, which, in Brazil, is in March. In principle, parents can choose which public school they will take their child to according to their preference. For instance, some parents - the more conscious of the importance of education - apply for schools that present high Ideb scores, [11] regardless of how far it is from their houses. It turns out that the schools have some criteria for the order of priority to accept the child. In addition to the proximity between the student's residence and the institution, having siblings in the school where the application is made is also a factor that the schools take into consideration to make the admission decision (for a thorough discussion on distance between school and home, and the endogeneity of the school choice decision by parents, see Card (1999, 2001)). This is due to the scarcity of places for all children in the primary school in the public system in Brazil, which causes enormous queues outside the schools during the enrolment period. Some parents have even to spend the night before the first day of registrations in the school street so as to try to guarantee a place for their children, although queuing does not assure that they will find a place in that particular institution.

When it comes to school entrance policy, usually it implies having children who are at different ages in the same class or grade, which may entail some cognitive consequences in the long run (Crawford et al., 2010; Dobkin and Ferreira, 2010; Black et al., 2011;

[11] Ideb is the main quality indicator of basic schooling in Brazil. Ranging from 0 to 10, it synthesises two concepts: the schools' pass rates and the learning level on Mathematics and Portuguese gauged by the *Prova Brasil*, the national examination for students from primary to high school, including adult education. The Ideb calculation is represented as follows:

$$\text{Ideb}_{ji} = N_{ij} * P_{ij}; 0 \leq N_j \leq 10; 0_j \leq 1; \tag{2.1}$$

where, $i = Ideb$ and School Census year; N_{ij} is equal to the average of Portuguese and Maths normalised average for a indicator between.

Peña, 2017). To mitigate this concern, the law 11,274/2006 establishes a cut-off date to enrol at the primary school, which is 6 years old completed or to be completed by the start of the school year. Hence, the new classes that will receive 6-year-old pupils do not enrol children with higher ages, except the students who repeated the first year, i.e., coming from a previous cohort. Furthermore, those who started their 8-year primary school - the previous cohorts - will not benefit from the 9-year elementary school. Likewise, six-year-olds who can read and write cannot be directly enrolled at the second year of the nine-year primary school since increasing the compulsory years of schooling objectives to extend the time and quality of the teaching-learning, instead of anticipating its conclusion. Additionally, both eight-year and nine-year curricula should coexist, during the former's process of extinction and the latter's implementation/transition phase.

The new nomenclature suggested by the National Education Council (CNE) (*Conselho Nacional de Educação)* for child education is displayed in the Table 2.6 . For more details on the Brazilian political-legal education system reforms, please see the Appendix.

Table 2.6: Primary school nomenclature by the National Education Council

School level	**Suitable starting age**	**Duration**
Child education	Until 5 years old	
Nursery	Until 3 years old	
Kindergarten	4 and 5 years old	
Primary School		9 years
Initial years	From 6 to 10 years old	5 years
Final years	From 11 to 14 years old	4 years

Note: Table elaborated with data from Basic Education Secretariat - Ministry of Education of Brazil SEB/MEC (2006).

The Resolution CNE/CEB No.3/2005 spells out that children who are up to 5 years old in the beginning of the school year must be enrolled at the kindergarten. Furthermore, the child education stage will be able to take care of children under 6 years old in two situations: (i) until the education system extends the Primary School to nine years of

duration respecting the deadline (in 2010) set by the law; and (ii) to reach all children who turn 6 after the date set for entry into schooling.

Regarding the municipalities, initially, they must consider whether they are linked to the state educational system rules or if they had their own system. In the former case, the municipality must comply with the State Council of Education (CEE) (*Conselho Estadual de Educação*) and submit its reform proposal for the Elementary School to analysis and approval of the CEE. The municipality with its own education system must just observe the norms already updated by its respective Municipal Council of Education (CME) (*Conselho Municipal de Educação*).

Automatic promotion

The Law 9,496 of 1996 introduced automatic grade promotion as optional in that year, with public state schools starting to adopt it in large numbers by the early 2000s.[12] Each educational system decided whether it was convenient to establish it in its premises. The organization of the public education system in Brazil is based on boosting the quality of teaching and learning in all schools, equality and equity for all students, access and permanence in all basic education. This expected improvement in the quality of educational services depends on the joint effort of its professionals for the implementation of public policies of education and articulation between the State Department of Education, Teaching Offices and schools of the state public education network. Each state or municipal network is entitled to opt for non-retention or not, and the schools linked to these systems should follow their guidelines accordingly. For instance, the State of São Paulo has adopted in its public education system the organization of elementary education in cycles, obligatory for the state public network. The Resolution SE 53 of October 2^{nd}, 2014 provides orientation for the re-organisation of the Elementary School in Continuing Pro-

[12]This section is based on information gathered by electronic contact to the Secretariat of Education of São Paulo, which explained the dynamic of the automatic promotion in that state, which I am roughly generalising here to all other states that adopt it.

gression Regime and on the Mechanisms of school support to the Elementary and Middle School students of the state schools, creating 3 learning phases, with a duration of 3 years each. The division of teaching in three cycles of learning aims to provide pedagogical conditions for children and adolescents to obtain more opportunities to be effectively served in their needs, enabling them to adapt learning times to their individual characteristics. The learning cycles, understood as time spaces interdependent and articulated among themselves, are defined throughout the nine years of Elementary Education as follows: i) Literacy Cycle, from 1^{st} to 3^{rd} year; ii) Intermediate Cycle, from the 4^{th} to the 6^{th} year; and iii) Final Cycle, from the 7^{th} to the 9^{th} year.

Failure in primary education has increased over the years, hence, to overcome the so-called "pedagogy of repetition" and exclusion, and to establish a pedagogy of human promotion and inclusion, the government decided to make it optional for the educational systems throughout the country to opt for continuous progression in elementary education, acknowledging its contribution to enhance the quality of teaching and learning. This strategy is an effort to fulfill the second MDG of universalising the basic education, to guarantee the access and permanence of the children at school and to regularise the flow of students in consonance with the relation age-grades.By adopting the system of continuous progression, the policy maker recognizes that the cycle exceeds the learning time in a single school year, stimulating less truancy and considering the students' idiosyncrasies. As a result, the scope of the students evaluations become an indicator of monitoring the learning progress, instead of being a mere instrument of pupils approval or disapproval.

As mentioned above, the public schools do not make the decision to adopt or not continuous progression (or automatic promotion); the choice is made by the municipal or state network to which the school is linked. As such, as my study is at the school level, rather than the municipal or state systems, I can avoid endogeneity concerns due to this specific issue.

2.4 Research Design

I focus on the schools' performance in the national examination *Prova Brasil*, pass and dropout rates, before and after changes in compulsory school entrance age[13] and automatic promotion laws. First, I examine whether schools who opted to enrol children aged six years old - instead of 7 years old, as it used to be before the new law 11,274/2006 - in the 1^{st} grade during the optional phase (2006-2010) of the reduction of school entrance age policy presented better results in Maths and Portuguese tests, higher pass rates and lower dropout rates. Second, I investigate if the schools who adopted automatic promotion in, at least, one year from 2011 to 2015 showed increased results regarding these same dependent variables.

One may question the possibility of self-selection bias stemming from the fact that students are not obliged to participate in the exam, although the schools are oriented to encourage and ensure that everyone is present on the day of assessment, including the ones with disabilities. One could also argue that the more responsible and hard working pupils would be more likely to be present in the day of the exams, or that the students may be given an incentive by the head of the school in terms of marks in their own school examinations if they take the *Prova Brasil* tests. Notwithstanding these valid concerns, it does not seem to be a serious problem since the national examinations are constructed to assess the schools, not the students themselves. Regarding the potential existence of a reward, in case it really happens, it would be offered to all students, thus the endogeneity problem here could be mitigated under this aspect.

Moreover, to avoid upward bias brought about by the self-selection of parents choosing where to enrol their children, I control for the mother's level of education and for the

[13] I analyse its first optional phase as a quasi-experiment.

proportion of parents that are interested in their children education.[14] Hence, I created these variables from the socio-economic questionnaire the pupils fill in during the *Prova Brasil* tests. My assumption is that the more educated she is and/or the more involved in the child education the parents are, the larger the chances of their decision be guided by their preference for better education.

2.4.1 Identification Strategy

Following the literature on education reform policies (Marques da Costa, 2018; Machin and Wilson, 2005; Eyles and Machin, 2015; Koppensteiner, 2014; Fertig and Kluve, 2005; Herbst and Strawiński, 2016; Carvalho, 2014; Okurut, 2018), to avoid endogeneity and biased estimates stemming from schools self-selecting to adopt the policies, I use Propensity Score Matching with Difference-in-Difference (PSM-DID) technique (Lechner, 2002; Leuven and Sianesi, 2003). The matching routines are able to remove that particular bias (to some extent) given that it creates a valid control group based on the differences between the schools that adopted the policies and the schools that did not.[15]

My study diverges from the aforementioned as I analyse Brazilian public schools, which has quite different institutions and school systems from England and Poland, countries considered in those empirical works. Additionally, I consider both reforms in my models; controlling for automatic promotion in the regressions whose main treatment is the school age reform. This study's purpose is to estimate the average treatment effect on the treated (ATT) for each reform so as to appraise the difference in public schools' performance at the *Prova Brasil* examinations, dropout and pass rates between schools which have adopted one or both of the treatments and equivalent schools which have not opted

[14] In the appendix, I list 4 questions that the students answer regarding their parents or legal tutors interest in their education: parents' incentive to attend the classes; parents' incentive to do the homework; parents' incentive to read and parents' incentive to study.

[15] In the methodology section, I show the limitations of the PSM technique concerning the unobserved characteristics that change over time.

for the treatments. Furthermore, I am able to evaluate not only the static results of the policies implementation, but also their dynamic effects by looking at two periods (t and $t+1$) and two cohorts. The latter analysis can just be done for the school admission age policy; the automatic promotion treatment does not provide me with enough data to assess more than one cohort.

The causal effect of lower school starting age and automatic promotion on the schools' performance at time $t+s$ can be identified as:

$$Y^1_{i(t+s)} - Y^0_{i(t+s)} \tag{2.2}$$

where I consider y_{it} as school i's scores in the *Prova Brasil*, dropout or pass rates at time t; and $Y_{i(t+s)}$ is the scores in the *Prova Brasil*, dropout or pass rates s period(s) later, with $s \geq 0$. The superscripts stand for the school decision: 1 if a school adopts the policy at t and 0 otherwise. Consequently, $Y^0_{i(t+s)}$ is the school i performance at period $t+s$ had it not opted for the new policies since period t.

Assignment to treatment

Although the ideal set to assess the causal relationship between the treatment and the response variables - the treatment effect - are randomised trials, most empirical studies rely on quasi-experiments to create the counterfactual group. My study sheds light on the optional phase of the school entrance age law, therefore I use matching methods to emulate a randomised experiment and to reduce the self-selection bias coming from the decision of the schools to adopt the new entrance age before the compulsory phase of the new policy. Since I cannot observe what would have happened had the treated schools not opted to lower the school age during 2006-2010, I have to create a relevant counterfactual to assess the causal effect from the difference between the outcomes of schools not exposed

to the treatment (counterfactual) and the schools actually exposed to the treatment (the factual). In this research, the outcomes of interest are (i) educational attainment based on the *Prova Brasil* scores; (ii) dropout rate; and (iii) pass rate. The treatment is school entrance age at 6 years old rather than at 7 years old; and automatic promotion (non-retention).

The main challenge to define the counterfactual is due to the fact that only one of the two aftermaths of Equation 2.2 is actually observed. For instance, if the school i decides to comply with the policy, its outcome variable will be y^1; the y^0 (the counterfactual result) cannot happen. As a result, I am not able to estimate the direct treatment effect for the treated schools. Then, I use the population average treatment effect (ATE) as follows:

$$ATE = E[Y^1_{i(t+s)} - Y^0_{i(t+s)}] \tag{2.3}$$

i.e. the difference in the expected responses of treated and non-treated.

Propensity Score Matching

I use PSM approach to select a valid control group from the untreated schools and to keep in my sample just the control schools with similar observed characteristics from the pre-policy period to the treatment group's (Becker et al., 2002); hence, minimizing the omission bias that my estimations would incur had I not matched on the observables with the PSM (Rosenbaum and Rubin, 1985; Heckman et al., 1998). After matching, I drop the schools from the control group that are not in the support group and estimate the ATT just with the schools from the control group and the treated schools. Thus, the ultimate matching goal is to associate each treated school with a school that has never adopted the policy given a set of observable characteristics.

I capture the difference between schools' performance after a school opts for the educational policy, by creating a dummy variable $TREAT_{it}$ (standing for beginning of the policy), which equals 1 if the school starts to adopt the policy at period t, and 0 otherwise. The average effect on treated schools' (the schools that lowered the entrance age to 6 years old or the schools that adopted non-retention) marks in the national examinations and on their dropout and pass rates is the ATT, which is expressed as follows:

$$\begin{aligned} ATT &= E[Y^1_{i(t+s)} - Y^0_{i(t+s)} \big| TREAT_{it} = 1] \\ &= E[Y^1_{i(t+s)} \big| TREAT_{it} = 1] - E[Y^0_{i(t+s)} \big| TREAT_{it} = 1] \end{aligned} \tag{2.4}$$

Similarly, the counterfactual is denoted by $E[Y^0_{i(t+s)} \big| TREAT_{it} = 1]$, but it is not observed, as explained above. Therefore, the average of the dependent variable regarding the control group (the schools that did not adopted the lower school entrance age or the automatic promotion) is a good substitute for the unobserved counterfactual (Heckman et al., 1998). It follows that Equation 2.4 is then:

$$ATT = E[Y^1_{i(t+s)} \big| TREAT_{it} = 1] - E[Y^0_{i(t+s)} \big| TREAT_{it} = 0] \tag{2.5}$$

Firstly, I conduct the PSM, which estimates the probability that a school adopts the policy in time t, based on their observable characteristics. To estimate the likelihood I use probit model as follows:

$$\begin{aligned} Pr\big(TREAT_{it} = 1\big) = \psi\big(&gender_{i(t-1)}, age_{i(t-1)}, presch_{i(t-1)}, race_{i(t-1)}, work_{i(t-1)}, failed_{i(t-1)}, \\ &liveparents_{i(t-1)}, mothsch_{i(t-1)}, parentsinterest_{i(t-1)}, pcs_{i(t-1)}, internet_{i(t-1)}, labs_{i(t-1)}, \\ &teachersroom_{i(t-1)}, teacherSdiploma_{i(t-1)}, D_t\big) \end{aligned} \tag{2.6}$$

where Pr denotes the predicted probability of school i adopting the policy at time t; and $\phi(.)$ is the normal cumulative distribution function; and D_t is a time dummy to capture time effects. Some *status quo* school's characteristics are inputted into the model

to estimate the propensity scores. The inclusion of the baseline characteristics in the probit model is to account for initial conditions, which may influence the ensuing changes in outcome or assignment to the treatment. Hence, ignoring the effects of initial conditions may bias the DID estimates, by wrongly establishing a causal relation between the post adoption of the policy and the outcome variables, when, in fact, the schools could have been enhancing their performance prior to the treatment, and the estimates may be just capturing that trend (Khandker et al., 2009).

To include the initial conditions, I can either lag the variables and insert them differentiated into the probability model - as I did in my main results -, or I can regress the dummy variable for the treatment with the 2007[16] exogenous variables to obtain the propensity scores from the baseline data (see section Sensitivity checks). As such, I choose to follow De Loecker (2007) and lag the variables in respect to the baseline year 2007. This procedure uses the PSM as the estimator to calculate the ATT via non-parametric analysis, whilst, the method of DID-FE uses parametric estimation to estimate the coefficient of treatment. As a result, both results have to be similar to each other, mainly in the signs and directions, as I report in the next sections, which is also a robustness check exercise. After estimating the scores, the next procedure is to check if the PSM is balanced within treated and control groups.

Following Garrido et al. (2014) and Abadie et al. (2004), I divide the sample in n intervals of the same size of the PSM and test for each interval if the score average for the treated corresponds to control. I keep re-dividing the original interval into more blocks and test it again, doing so as many times as necessary to achieve the equivalence between control and treatment groups in all intervals. Eventually, I can test the balance of each confounder, under the null hypothesis that the covariates do not need to be changed, i.e., it is not necessary to re-define the variables included in the probit model, by, for example,

[16]The baseline year for the school entrance age policy is 2007; and for the automatic promotion policy is 2009.

inserting higher order terms or splines to increment the accuracy of the balance. If the null is rejected, then the modification of the covariates is recommended to reach a more accurate balance of scores (Blundell and Costa Dias, 2000).

Using the balanced scores to estimate the probability of starting to adopt the policy, the treated schools are finally matched with their counterparts from the control group. Matching mechanisms abound in the literature, for instance, one-to-one; caliper Cochran and Rubin (1973); nearest neighbour (nn); radius; kernel Heckman et al. (1997, 1998); stratification and Mahalanobis matching Rubin (1980). The latter is useful, in particular, for multiple treatment framework. Caliendo and Kopeinig (2008) explain that any of these matching algorithms may be applied to build the counterfactual. The main difference among them is the way they determine the acceptable neighbourhood - or how the similarity between control and treated is weighted - for the control group to be classified and captured as on common support, thus being considered a match. It follows that each of these approaches present pros and cons that have to be evaluated by the researcher when picking one of them (for more details on the techniques, see (Sianesi, 2001)).

A major issue in non-parametric analysis is the trade-off between bias and variability. For example, in the Kernel matching, each treated school receives a weight and their control counterparts are weighted in conformity with the distance from the treated school, given a certain bandwidth. Hence, the choice of this bandwidth is crucial to the matching procedure, which will be defined by the researcher according to her preference to bias/variance. Thus, on one hand, the larger the chosen range is, the higher the bias will be. On the other hand, a more even density function can be estimated by the kernel approach, representing a more accurate PSM fit, for the reasons given by Garrido et al. (2014) and Caliendo and Kopeinig (2008). In using the Kernel approach, I can mitigate the weight of the contribution to explain the response variable given by the less similar control schools to the treated schools, within a chosen bandwidth of the PSM distribu-

tion. The Kernel estimator smooths the matching curve for the PSM estimation, reducing the median standardised bias between treated and reference groups, and it uses as much information as feasible to do the matching, which make the estimator more precise than the others aforementioned, without losing quality (Sianesi, 2001; Caliendo and Kopeinig, 2008). For this work, as the sample size is large, and as I prefer less bias in my estimates, my bandwidth is 0.01.

The use of PSM-DID is able to mitigate the effect of unobservables that are the same over time. Thus, as long as the researcher controls for school and time fixed effects, she will have to deal just with the confounders that change along the period of the panel, which will be absorbed by the error term. Not being capable of correcting the bias caused by the time-varying unobservables is a limitation of the PSM-DID routine Blundell and Costa Dias (2000), whose PSM-DID estimator is shown below:

$$ATT_{PSM-DID} = \frac{1}{N_T} \sum_{i \in T} [\Delta Y_{i(t+s)} - \sum_{j \in C} w_{ij} \Delta Y_{j(t+s)}] \tag{2.7}$$

where N_T is the number of schools who adopt the new policy in the common support, ΔY_i is the difference between the Mean of the performance of schools before and after adopt the program of school i from the intervention group; ΔY_j denotes the the pre and post difference of school j from the comparison group and w_{ij} is the weight for control school j in the construction of the estimated expected counterfactual result for the treated school i, determined by the score, $\sum_{j \in C} w_{ij} = 1$. (De Loecker, 2007).

The most critical and strong identifying assumption to be met within the DID is the so-called Parallel Trend Assumption (PTA), which needs to hold to guarantee the internal validity of the DID technique. To fulfill this requirement, the difference between the treated and control schools must be constant over time, that is, if by graphic or summary statistics inspection[17], the researcher spots non-parallel trajectories of both control and

[17]Tables A6, A7, A8, A9 show that most of the covariates included in the matching present non-

treatment groups, then it suggests that there are unobservables that vary throughout the panel period that have not been controlled for. That would be the case, for instance, if selection for treatment is influenced by individual-transitory shocks on past outcomes (Ashenfelter's dip[18]) as explained by Abadie (2005).

As the PSM aims solely at removing the bias originated from constant unobservables, as a result, even after the matching procedure, the PTA may be violated as the econometrist cannot control for the former. Therefore, once disregarded the PTA, the estimation of the causality stemming from the treatment effect to the outcome variables will be severely biased (for more details on the PTA, see, e.g., (Angrist and Pischke, 2008; Imbens and Wooldridge, 2009; Card and Krueger, 1993)).

Propensity Score Matching Diagnostic

To assess the rigour of the balancing of the matching applied in this chapter, I follow the tests suggested by the literature on Propensity Score Matching (see e.g. (Austin, 2009; Caliendo and Kopeinig, 2008; Rosenbaum and Rubin, 1985; Blundell and Costa Dias, 2000; Garrido et al., 2014; Sianesi, 2001; Abadie et al., 2004)): i) comparison between the means of the variables inputted in the matching before and after it; ii) observation of the standardized bias for all these variables used to perform the PSM.

The first test consists of comparing the means before and after matching on observables from the treatment and control groups. The expected result for this test is that, after matching, the difference in means between treatment and control groups be statistically equal to zero, i.e, the difference between the two groups should drop substantially after the matching since it is the main purpose of the PSM. The test uses the t-test to

significant difference between treated and control groups after the matching.

[18]"Ashenfelter's dip" is known as the empirical pattern that the Mean earnings of participants in employment and training programs generally decline during the period just before the program (Smith, 1996). "Dip" here Means "trend".

assess the significance of the difference between the means of each variable. It turns out that the customary t-test may lead the researcher to dubious conclusions as Austin (2009) and Garrido et al. (2014) point out. It occurs because the t-test hinges on the sample size, hence, if the number of observations is too high, as it is the case of my sample, the test may indicate non-significance even though the balance of the confounders is not accurate. In this manner, the literature recommends that the second test be used to form the final opinion on the balancing *status* of the covariates.

For the second test, I expect to find the smallest standardised bias suggesting that my matching is balanced, thus, I can use the matched control group to compare with the treated schools. The normalised difference for a continuous variable is:

$$\underbrace{Norm - Diff}_{\text{standardized difference}} = 100 * \frac{\bar{X}_{Tr} - \bar{X}_{Co}}{\sqrt{\frac{S^2_{Tr} + S^2_{Co}}{2}}} \tag{2.8}$$

where $\bar{X}_{Co}$ and $\bar{X}_{Tr}$ are the Mean of each control variable included in the matching for the control and the treated groups; and S^2_{Tr} and S^2_{Co} represent the sample variances for the treated and control schools.

For a categorical variable, the standardised difference is obtained by:

$$\underbrace{Norm - Diff}_{\text{standardized difference}} = 100 * \frac{\hat{J}_{Tr} - \hat{J}_{Co}}{\sqrt{\frac{\hat{J}_{Tr} * (1 - \hat{J}_{Co}) + \hat{J}_{Co} * (1 - \hat{J}_{Co})}{2}}} \tag{2.9}$$

where $\hat{J}_{Tr}$ and $\hat{J}_{Co}$ stand for the Mean of the categorical variable J in treated and control groups.

If the standardised difference between the matched and the unmatched sample is small - under 20% - then the balance achieved is considered good. That being the case, the matched control group can be used to perform the DID technique assuming that group the comparison one to the treated schools. I employ the figure 20% indicated by the seminal work (Rosenbaum and Rubin, 1985), and as the approach used by other researchers

when using PSM, such as Sianesi (2001); Wooldridge (2010); Garrido et al. (2014), among others. The balance tests and the visual balance identification are in Tables A6, A7, A8, A9.

Another step emphasised by Sianesi (2001) when conducting the PSM balancing tests is to check the pseudo-R^2 [19]. I then compare the pseudo-R^2 before and after matching expecting to find a pseudo-R^2 after the PSM smaller than the one found before the procedure. Particularly, a desirable lower pseudo-R^2 will express that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups. Figure 2.1 depicts the observations matched with stars and unmatched with black balls of a PSM execute with the sub-set of municipal schools regarding the treatment automatic promotion. The graph shows that the matched observations are quite close to the vertical line in the zero percent standardised bias, which means that there is very little difference between treatment and control groups after the matching procedure. Furthermore, this plot is representative of a decrease of the pseudo-R^2 from 0.167 before matching to 0.004 after matching, which is the expected result for a good matching.

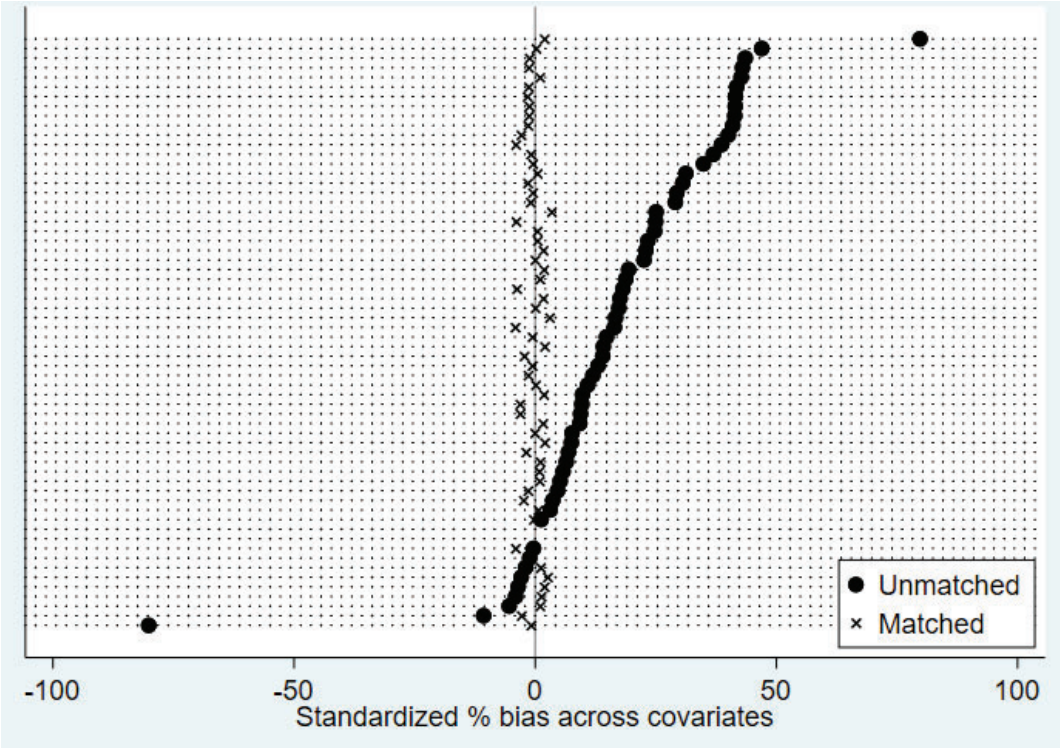
Unobserved Heterogeneity

If the unobserved school heterogeneity persists over time, overlooking it will lead to a serious estimation bias as it would violate the common trend assumption. For instance, a school displayed a particular performance in the *Prova Brasil* examinations in a given year which is due to the influence of a certain unobserved and pre-existent heterogeneity,

[19] The pseudo-R^2 indicates the overall fit of the model. It captures the improvement in the fit of the Maximum likelihood function used for the estimations, when compared to a model that includes just the constant in the regression.

$$Pseudo - R^2 = 1 - \frac{\text{ln likelihood(estimated model)}}{\text{ln likelihood(model with just an intercept)}} \tag{2.10}$$

Figure 2.1: PSM balancing using only municipal schools (primary cycles)



Note: The graph shows that the matched observations are quite close to the vertical line in the zero percent standardised bias, which means that there is very little difference between treatment and control groups after the matching procedure. The pseudo R-square before the matching was 0.167, and after matching, 0.004. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

such as psychological assistance to the families of students who have poor records. Another example related to the case of automatic promotion policy is if schools with high performing pupils and low retention rates adopt non-retention policy, then the response variables and the policy are correlated not only because of the treatment, but also for those other factors, breaching the common trend assumption. If that is the case, the results would be overestimating the actual effects of the treatment on the response variable - Maths and Portuguese scores; pass and dropout rates. In this sense, if I employ standard econometric approaches that do not take into account unobserved characteristics (that do not change over time), I will be neglecting the real effect of the policy (Koppensteiner, 2010).

Therefore, I consider that the error term ν_{it} in Equation 2.11 consists of two compo-

nents, the unobserved school-specific time-invariant effects, k_i, and an exogenous random disturbance, μ_{it}, which gives:

$$\nu_{it} = k_i + \mu_{it} \tag{2.11}$$

where k_i is normally distributed and independent on μ_{it}, and the latter is normally distributed and independent of the independent variables $\forall$ i and t. The unobserved school idiosyncrasies can be controlled either by Fixed Effects (FE) or Random Effects (RE) technique. The former estimates the unobserved characteristics as a parameter, and the latter assumes that it is a random variable. Nonetheless, following Wooldridge (2010) and Vanino (2016), I opted to use the random effects approach when estimating the probability of the schools to adopt the policy since treating the unobserved heterogeneity as a parameter may cause biased estimations of the β_i and of the response variable as a consequence of the incidental parameters problem.

2.5 Empirical Findings

I divide this section into two: the first one presents the results for the compulsory school entrance age policy and the second one covers the findings for the automatic promotion policy.

2.5.1 Compulsory school entrance age

Table 2.7 presents the main estimates of the effect of the adoption of the new compulsory school entrance age on the national examination scores *Prova Brasil*, dropout and pass rates at the school level, for the whole sample. The estimated coefficients refer to the ATT captured by the PSM kernel with bandwidth 0.01. I also conducted the PSM using one to one, nearest neighbour and caliper matching, whose results can be found in the Table A3. The Table presents results for two cohorts. First, cohort 1, in the time $S = 0$ comprises the students who enrolled at the first grade at age 6 in 2007 and took the

exams in 2011 in the fifth grade; cohort 1 in the time $S = 1$ refers roughly[20] to the same students, but who passed the tests in 2015 in the ninth grade. Second, I look at cohort 2, which encompasses the children who started school at the first grade with 6 years old in 2009 and took the fifth grade tests in 2013. For the latter cohort, I cannot examine the second period, $S = 1$ because there is no data available for the exams in 2017, year when they might have taken the exams in the ninth grade. This Table is also divided into set of columns named after letters: letter (a) brings results for Portuguese scores; (b) for Maths; (c) for dropout rates and (d) for pass rates.

From the first set of columns (a), one can see that overall both cohorts increased their Portuguese scores. Cohort 1's raised by 0.3 standard deviation (SD) in $S = 0$ (6.70 marks) and in $S = 1$ (4.79 marks). The second cohort saw an increase of 0.41 SD (10.60 marks). The set of columns (b) shows the scores for Maths. Cohort 1 did not have any statistically significant change in its scores in $S = 0$, but showed an increase of around 0.14 SD (2.41 marks). Cohort 2 shows a rise of 0.29 SD (10.35 marks). Regarding the pass rates, none of them observed significant changes. When it comes to dropout rate, one would expect a decrease in the figures, although what can be noticed is that both cohort 1 in time $S = 1$ and cohort 2 had an increase in the dropout rates: 1.46 pp and 2.61 pp respectively. The small magnitudes of the coefficients are very common to be found in Economics of Education works. For example, Herbst and Strawiński (2016) found a change of 0.11 and 0.04 SD in writing and Maths for the Polish students after the age to enrol at the first grade lowered from 7 to 6. Looking at pupils in Tlaxcal, Mexico, Peña (2017) registered results around 0.3 SD. These works found a decrease in Maths and reading/writing scores when comparing younger children in the same classroom of older ones. My work adds to the literature insofar as it investigates the performance of students in the same classroom with the same age of entrance. It means that I am not comparing the different ages within the

[20]By roughly, I mean the imprecision of the components of the cohort. That is, some students may have failed, changed school and even dropout between 2011 and 2015; hence, it is not necessarily all the same students who took the examinations in 2011, the ones who took in 2015.

Table 2.7: Whole sample - Treatment: compulsory school entrance age - (Kernel)

Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese				(b) Maths			(c) Dropout			(d) Pass		
ATT	6.70***	4.79***	10.60**	3.01	2.41*	10.35***	0.55	1.46*	2.61**	-0.55	1.02	-0.41
	[1.82]	[1.60]	[4.83]	[2.40]	[1.34]	[2.60]	[1.13]	[0.88]	[1.22]	[1.36]	[1.48]	[3.37]
N(control)	306	198	14	306	198	14	308	76	16	308	74	16
N(treat)	16,803	9,978	23,487	16,799	9,978	23,494	22,147	5,922	28,160	22,131	5,910	28,153
N(total)	17,109	10,176	23,501	17,105	10,176	23,508	22,455	5,998	28,176	22,439	5,984	28,169

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

same grade, rather I am looking at similar cohorts (same entrance age) at the school level.

When I split the sample into state and municipal schools (see Tables 2.8 and 2.9), Portuguese scores remain pretty much the same for both types. Whereas, municipal schools are affected by the reform in Maths, pass (a decrease of 4.58 pp) and dropout rates (an increase of 3.96 pp). These results may inform that the change in the compulsory schooling start age to 6 years old made more difficult to students to pass, which can be a good sign that the policy is indeed transforming, even slightly, the quality of education in public schools. Additionally, there is literature on compulsory school start age stressing the lower performance of younger students when compared to the older cohorts due to the lack of maturity and the challenge of starting reading and writing earlier (see the section Literature review for a discussion on this issue). As a result of the increase in the repetition rates, cohort 1 in time $S = 1$ showed a rise in the dropout rate of 3.96 pp. For instance, McEwan and Shapiro (2008) studied a case for Chile and found that a one-year postponement in school admission lowers the probability of failing the first grade by two percentage points; it also increases fourth and eighth grade test scores by more than 0.3 SD, with larger effects for boys[21].

It is worth noting that the pass and dropout rates are collected from the School Census data set, which is a set of information delivered by the schools. On the other hand, the Portuguese and Maths scores are information coming from the results in the *Prova Brasil*, rather than the schools internal examinations. In addition, the signs of Portuguese and Maths scores are positive, and the signs for pass rates are negative, which may appear to be counter-intuitive. It turns out that, the negative sign for pass rates are quite meaningful in accordance with the Brazilian institutions and educational system as the students might be failing the grades because they may be presenting internal tests scores lower than the scores they achieve at the *Prova Brasil*. In other words, the *Prova Brasil* may

[21]For more empirical evidence in support of the magnitude and the signs of the coefficients found in my study, please, see the Literature review section.

Table 2.8: Only municipal schools - Treatment: compulsory school entrance age - (Kernel)

Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese				(b) Maths			(c) Dropout			(d) Pass		
ATT	3.43*	3.60*	11.86**	0.84	1.06	12.78***	1.29	3.96***	1.28	-4.58	1.20	1.1
	[1.96]	[2.03]	[6.09]	[2.21]	[1.62]	[4.14]	[0.94]	[1.68]	[1.16]	[1.37]***	[2.79]	[4.65]
N(control)	155	147	9	155	147	9	157	45	11	157	43	11
N(treat)	26,921	6,449	18,363	26,913	6,449	18,374	32,920	3,463	20,984	32,882	3,434	20,975
N(total)	27,076	6,596	18,372	27,068	6,596	18,383	33,077	3,508	20,995	33,039	3,477	20,986

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate SE by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

be easier than the internal examinations at schools to show to the society that the government is tackling the low quality of education. My sample is from 2007 to 2015, period in which the left-wing party was in the power in Brazil, whose main flag was addressing poverty and education in the country. All in all, the fact that pass rates have decreased throughout the period is another indication that the policy itself is not able to increase the quality of education in Brazil.

Table 2.9: Only state schools - Treatment: compulsory school entrance age - (Kernel)

Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese				(b) Maths			(c) Dropout			(d) Pass		
ATT	5.70***	6.82**	2.26	2.03	2.39	-7.08	-0.29	-0.13	1.99	1.98	2.04	-3.06
	[1.67]	[3.32]	[9.03]	[2.21]	[3.53]	[6.40]	[1.12]	[1.13]	[3.51]	[1.32]	[2.61]	[10.11]
N(control)	275	51	8	275	51	8	277	31	8	277	31	8
N(treat)	6,054	3,497	7,688	6,054	3,497	7,686	8,565	2,238	9,838	8,557	2,255	9,835
N(total)	6,329	3,548	7,696	6,329	3,548	7,694	8,842	2,269	9,846	8,834	2,286	9,843

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

2.5.2 Automatic Promotion

Looking at the second policy analysed in this chapter - automatic promotion - Table 2.10 shows that, for the whole sample, the adoption of automatic promotion increased the Maths scores by 0.19 SD (4.38 marks) for the cohort who started school in 2009 and took the exams in 2013. For the remaining response variables, the reform did not generate any significant effect for those schools that opted for non-retention.

Table 2.10: Whole sample - Treatment: Automatic promotion - (Kernel)

Cohort 1	S=0	S=0	S=0	S=0
Dep. Var.	**(a) Portuguese**	**(b) Maths**	**(c) Dropout**	**(d) Pass**
ATT	1.4	4.38***	-0.1	0.34
	[1.14]	[1.57]	[0.11]	[0.42]
N(control)	4,968	4,968	9,452	9,448
N(treat)	844	844	992	992
N(total)	5,812	5,812	10,444	10,440

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Considering only state schools in Table 2.11, the dropout rate increased by 1.10pp and pass rate by 3.80pp. The latter result is quite intuitive as the policy aim was rising the pass rates to stimulate students not to dropout early from school motivated by failures. It turns out that it seems the policy did not achieve its objective as the dropout rates continued to soar. Moreover, when it comes to only municipal schools sample, Table 2.12 shows that only Maths results had a slight increase for the schools that adopted the non-retention policy (around 0.11 SD).

The literature on automatic promotion is not unanimous. For example, Koppensteiner (2014), investigating only the state of Minas Gerais (one of the wealthiest, located in the Southeast region, which is the better-off in Brazil), found a tiny negative effect of 0.07 SD of the policy on the Maths scores, and non-significant effect on the dropout rates. Concerning the Northeast of Brazil (one of the poorest regions in the country) Gomes-Neto

Table 2.11: Only State schools - Treatment: Automatic promotion - (Kernel)

Cohort 1	S=0	S=0	S=0	S=0
Dep. Var.	**(a) Portuguese**	**(b) Maths**	**(c) Dropout**	**(d) Pass**
ATT	1.34	5.65	1.10***	3.80***
	[9.44]	[16.11]	[0.30]	[1.77]
N(control)	765	765	2,951	2,951
N(treat)	140	140	144	144
N(total)	905	905	3,095	3,095

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

and Hanushek (1994) concluded that Maths scores improved more than Portuguese via grade repetition in that region's public schools. Hence, I can consider that my results are consistent with these empirical works, adding to the literature since I investigate all Brazilian states and municipal schools and found no evidence that passing all students in disregard of whether they are capable to go to the next grade affects the dropout rates and significantly increases literacy and/or the quality of education in Brazil. Other papers in this area show very little or non-significant results regarding the same response variables as mine, indicating that this policy is not the best approach to tackle quality of education and minimise dropout rates either (see for example, (Jacob and Lefgren, 2009; Manacorda, 2012)).

Table 2.12: Only municipal schools - Treatment: Automatic promotion - (Kernel)

Cohort 1	S=0	S=0	S=0	S=0
Dep. Var.	**(a) Portuguese**	**(b) Maths**	**(c) Dropout**	**(d) Pass**
ATT	0.82	2.61*	-0.2488	-0.21
	[0.72]	[1.56]	[0.12]	[0.46]
N(control)	3,931	3,931	6,128	6,124
N(treat)	704	704	848	848
N(total)	4,635	4,635	6,976	6,972

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works (I used 500 repetitions). I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM. Levels of significance: * 0.10; ** 0.05; *** 0.01.

2.6 Sensitivity Checks and Limitations

After checking the balance of my PSM, I can merge the "on support" group with the whole sample and run the DID-FE, as detailed in the next sections.

2.6.1 Difference-in-differences with Fixed Effects

The new sample with just on support control and treated schools allows me to proceed to the second part of my methodological strategy, which consists of applying DID-FE to capture the impact of the adoption of the lower school entrance age (or the automatic promotion) on the schools performance in the national examinations *Prova Brasil*, dropout and pass rates. The model of this section uses the DID estimator, and I also control for school fixed effects.

The DID is a quasi-experimental design consisting of, firstly, gauging the difference in the response variables - exams scores, pass and dropout rates - between the period when the schools did not adopt the policy and after they opted by the policy. By comparing the changes in outcomes over time between a population that is enrolled in a program and a population that is not, the DID estimation takes into account the past performance and the idiosyncrasies of the schools in the sample, avoiding potential bias in the estimations stemming from pre-policy characteristics. Secondly, one need to take the difference on the difference calculated for the treated schools minus the control schools. The aftermath is an estimation without the effects of common trends from before the policy adoption.

Following Koppensteiner (2010), under the above identifying assumption, I estimate the effect of the two educational policies discussed in this chapter on the schools' performance in the *Prova Brasil*, dropout and pass rates using the model in Equation 2.12:

$$Y_{st} = \beta_0 + d_s + d_t + \phi d_{st} + \beta_1 X_{st} + \epsilon_{st} \tag{2.12}$$

where Y_{st} is the test score, dropout or pass rate for school s at time t, d_s is a school fixed effects capturing school-specific time invariant effects, d_t is time fixed effects which capture the common time trend of control and treatment groups, d_{st} is the time/treatment interaction term containing information on the treatment status of schools, which varies over time. The coefficient of interest is ϕ, which informs the the average treatment effect of the adoption of the educational reform on the response variables examined here. X_{st} denotes a matrix of exogenous covariates for class and school characteristics - the same used in the PSM model. Therefore, if any difference that lead to the adoption of the policy is included in the school fixed effects, then the parallel trend assumption is not violated.

Bertrand et al. (2004) emphasise the importance of clustering the standard errors at the identification variable level to guarantee that the interaction coefficient is not overestimated. According to the authors, most DID studies rely on years of data and focus on serially correlated outcomes. Nonetheless, most papers disregard the bias in the estimated standard errors that serial correlation introduces. This raises concern since the explanatory variable of interest in DID estimation (e.g. the passage of law) is itself highly serially correlated, producing inflated bias in standard errors. Hence, following the authors, I clustered the DID-FE standard errors per school as Table 2.13 shows.

Table 2.13 depicts roughly the same magnitude and direction of signs as the main results using PSM-DID, corroborating the latter. As already explained in other sections, the DID-FE is a parametric estimation, while the PSM-DID is non-parametric, thus the magnitude difference between the two estimations.

Table 2.13: DID-FE - Treatment: compulsory school entrance age (World Bank Method)

	Whole sample			**State schools**		**Municipal schools**	
	Cohort 1		Cohort 2	Cohort 1		Cohort 1	
Dep.Var./Period	S=0	S=1	S=0	S=0	S=1	S=0	S=1
(a) Portuguese							
Treat(interaction)	3.99***	4.02	25.81	4.23***	9.22*	6.13*	1.88
	[1.14]	[3.45]	[21.13]	[1.28]	[5.41]	[3.82]	[4.71]
N(Obs)	19,266	10,793	16,063	7,525	5,259	11,698	5,348
(b) Maths							
Treat(interaction)	0.84	0.44	19.17	1.41	3.64	0.91	-1.1
	[1.52]	[3.16]	[47.94]	[1.72]	[4.88]	[7.52]	[4.38]
N(Obs)	19,264	10,793	16,067	7,525	5,259	11,696	5,348
(c) Dropout							
Treat(interaction)	0.05	-0.82	6.04**	-0.13	-0.72	1.80***	-0.24
	[0.76]	[1.26]	[2.43]	[0.85]	[1.84]	[0.63]	[1.79]
N(Obs)	25,340	11,366	21,219	10,479	5,493	14,812	5,678
(d) Pass							
Treat(interaction)	1.12	3.45*	-4.14	3.26***	2.14	-3.34	3.65
	[0.92]	[1.82]	[6.57]	[1.00]	[2.71]	[1.69]**	[2.55]
N(Obs)	25,305	11,297	21,217	10,473	5,470	14,804	5,635
School Fixed Effects	yes	yes	yes	yes	yes	yes	yes
Year dummy	yes	yes	yes	yes	yes	yes	yes
Students controls	yes	yes	yes	yes	yes	yes	yes

Notes: Estimations clustered by school, producing robust standard errors. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table 2.14: Treatment: Primary cycles DID-FE (World Bank Method)

	Whole sample	**State schools**	**Municipal schools**
Dependent Variable/Period	S=0	S=0	S=0
(a) Portuguese			
Treat(interaction)	-0.85	-3.04	-0.45
	[0.73]	[3.60]	[0.74]
N(Obs)	8,770	920	7,459
R-squared	0.2110	0.0779	0.2464
(b) Maths			
Treat(interaction)	-2.49	1.64	-2.76
	[1.44]*	[6.20]	[1.52]*
N(Obs)	8,770	921	7,458
R-squared	0.0316	0.0231	0.0479
(c) Dropout			
Treat(interaction)	-2.49	0.31*	-0.18
	[1.44]*	[0.16]	[0.10]*
N(Obs)	8,770	2,752	10,634
R-squared	0.0268	0.0032	0.0355
(d) Pass			
Treat(interaction)	0.73*	-0.72	-1.51
	[0.39]	0.56	[0.26]***
N(Obs)	14,393	2,752	10,634
R-squared	0.0608	0.0050	0.0840
School Fixed Effects	yes	yes	yes
Year dummy	yes	yes	yes
Students controls	yes	yes	yes

Source: Estimations clustered by school, producing robust standard errors. Levels of significance: * 0.10; ** 0.05; *** 0.01.

2.6.2 Other PSM techiques

The main results in this chapter are based on the PSM-DID technique using the Kernel matching with bandwidth 0.01 for the reasons exposed in section 4. In addition to the Kernel type, I also ran Nearest Neighbour, Caliper (0.01) and One to one matching as robustness checks[22]. From the Tables A3, A4, A5 A10, A10 and A12, it is noticeable that the results are consistent with the main Kernel results, across the various specifications.

2.6.3 Placebo Tests

In this section, I perform the PSM-DID estimations based on placebo treatments. From the group of non-treated schools, I draw a random sample of schools - choosing a 10% proportion of the total sample - that I then considered to be the schools that adopted the new school entrance age (or automatic promotion) and implemented the PSM-DID estimator to compare the schools' performance in the *Prova Brasil* exams, the dropout and pass rates to the matched control schools (Huber et al., 2013). I simulated this process 1,000 times and found no significant impacts on the dependent variables (Portuguese and Mathematics scores, dropout and pass rates) caused by the placebo treatment schools as Table 2.15 shows. Therefore, the non-significant results indicate that my main findings displayed in section 5 are not only statistically significant, but also robust.

Concerning the second policy in analysis in this chapter - automatic promotion - I conducted the same placebo treatment test as described above and also found no significant effects of that policy on the outcome variables caused by the placebo treatment (see Table 2.16). Despite the seemingly counter-intuitive significant and negative effect on pass rates, it makes sense as the real treated schools did not present any significant average effect on pass rates due to the policy.

[22]The explanations of each technique can be found in section 4 of this chapter.

Table 2.15: PSM-DID estimations using placebo treatment: compulsory school entrance age

Dep.Var	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1			Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0		S=0	S=1	S=0
	(a) Portuguese			**(b) Maths**			**(c) Dropout**				**(d) Pass**		
ATT	-2.49	-1.55		-3.69	3.9		-0.17	2.34		0	3.65	2.42	
Boostrap S.E.	[3.66]	[5.22]		[4.95]	[4.69]		[2.85]	[2.19]	[1.81]		[2.83]	[3.17]	

Notes: Standard Errors bootstrapped with 1,000 repetitions. I ran 1,000 PSM-DID simulations using placebo treatment. The placebo treatment is not significantly causing the outcome variables from letters a) to d) in the table.

Table 2.16: PSM-DID estimations using placebo treatment: automatic promotion

Cohort 1	S=0	S=0	S=0	S=0
	(a) Portuguese	**(b) Maths**	**(c) Dropout**	**(d) Pass**
ATT	0.77	1.21	0.08	-0.7***
Boostrap S.E.	[0.86]	[1.45]	[0.08]	[0.25]

Notes: Standard Errors bootstrapped with 1,000 repetitions. I ran 1,000 PSM-DID simulations using placebo treatment. The placebo treatment is not significantly causing the outcome variables from letters a) to d) in the table.

2.6.4 Limitations

This work does not follow an experimental research design as I analyse one policy which is totally optional (the automatic promotion); and although the second one (the compulsory school entrance age) is compulsory and reaches all the public and private schools, I look at the optional phase because there is not enough time lapsed since the beginning of the compulsory phase (in 2010), results and data set to analyse it. In spite of the fact that randomisation cannot be guaranteed, there are similarities between control and treated groups because I just kept public schools in my sample and controlled for a plethora of schools, students and staff features. Moreover, the remaining differences between the two groups were overcome by the PSM method, persisting just generally minimal and relatively stable over time differences. Hence, I can sustain a plausible causal inference between the treatments and the outcomes of my research (Murnane and Willett, 2010).

Even though the data set on education in Brazil is resourceful and detailed, there are some observed and unobserved variables and some observations missing from the different data sets that I merged to build the panel data for this chapter. That may influence the final estimations of student dropout and pass rates, as well as the scores. Thus, the estimated effect of the policies may justifiably attract some critical reviews.

2.7 Conclusion

In this chapter, I analyse the effect of two independent educational policies on the schools performance at the national examination *Prova Brasil*, dropout and pass rates. The first policy made it optional for schools to adopt the automatic promotion from the first to the ninth grades. In 2000, a massive number of schools started to introduce the non-retention policy. I found no evidence of the impact of this policy on the Portuguese and Maths' scores at the *Prova Brasil*. In addition, I found significant and positive coefficients for pass rates, which is the aim of the practice of non-retention. However, the results indicate also a positive relationship between non-repetition in the primary school and dropout rates, which may be signalling that the policy itself does not suffice to change the alarming dropout scenario and to keep students away from the streets.

The second policy increased the compulsory years of schooling from 8 to 9 years, by anticipating the school starting age from 7 to 6 years old. The law had an optional phase, from 2006-2010, offering the schools the opportunity to choose whether they would accept younger children in the first grade. This non-mandatory phase was due to the necessity of important changes in the infra-structure, staff training and didactic resources. From 2010 onwards, all the schools were obliged to welcome children at the age 6 in the first grade. My results indicate that this policy raised the Portuguese and Maths scores in the national examinations, but also increased the dropout rates and lowered the pass rates. The latter is likely to happen to do the fact that the schools internal examinations are harder them the *Prova Brasil* due to political issues discussed in the previous section.

All in all, these results suggest that it is not only by issuing new educational rules without attempting to the states, municipalities, schools and children idiosyncrasies that the country will enhance the quality of education. The stakeholders particularities of a vast Brazil must be taken into consideration; interdisciplinary and contextualised methodologies must ensure meaningful learning for students. Therefore, new profiles of staff, training

them, monitoring not only the students, but also the schools, and the professionals are of the utmost importance to improve the quality of education in Brazil. These remarks can be applied to other developing countries as well, with governments ensuring adequate availability of complementary tools to influence the accomplishment of the policies objectives, the efficiency and quality of primary education.

Future research can cast light on the compulsory SSA policy phase, when there is data set available to assess the causality of the reform on the response variables analysed in this paper. Furthermore, other causal relations can be considered; for example, crime in Brazil or in one specific state in Brazil and the SSA, via the incapacitating effect stemming from more years of schooling.

Appendix

Table A1: Marks scale for Portuguese exam in *Prova Brasil*

	Portuguese		
Level	5th grade	9th grade	high school
Up to level 1	0 - 149 marks	———	———
Level 1	———	200 - 224 marks	225 - 249 marks
Level 2	150 - 174 marks	225 - 249 marks	250 - 274 marks
Level 3	175 - 199 marks	250 - 274 marks	275 - 299 marks
Level 4	200 - 224 marks	275 - 299 marks	300 - 324 marks
Level 5	225 - 249 marks	300 - 324 marks	325 - 349 marks
Level 6	250 - 274 marks	325 - 349 marks	350 - 374 marks
Level 7	275 - 299 marks	350 - 374 marks	375 - 399 marks
Level 8	300 - 324 marks	375 - 400 marks	400 - 425 marks
Level 9	325 - 350 marks	———	———

Source: Brazilian Ministry of Education.SAEB (2006)

Table A2: Marks scale for Mathematics exam in *Prova Brasil*

Mathematics			
Level	5th grade	9th grade	high school
Level 1	125 - 149 marks	200 - 224 marks	225 - 249 marks
Level 2	150 - 174 marks	225 – 249 marks	250 - 274 marks
Level 3	175 - 199 marks	250 - 274 marks	275 - 299 marks
Level 4	200 - 224 marks	275 - 299 marks	300 - 324 marks
Level 5	225 - 249 marks	300 - 324 marks	325 - 349 marks
Level 6	250 - 274 marks	325 - 349 marks	350 - 374 marks
Level 7	275 - 299 marks	350 - 374 marks	375 - 399 marks
Level 8	300 - 324 marks	375 - 399 marks	400 - 424 marks
Level 9	325 - 350 marks	400 - 425 marks	425 - 449 marks
Level 10	————	————	450 - 475 marks

Source: Brazilian Ministry of Education. SAEB (2006)

Table A6: Matching diagnostic - treatment: compulsory school entrance age (Kernel method); cohort 1; s=0; only STATE SCHOOLS)

	Unmatched	Mean			Mean %reduct	t-test	$p > \|t\|$
Variable	Matched	Treated	Control	%bias	bias		
dpropfemale	U	0.92831	1.9712	-13.9		-2.15	0.032
	M	0.92831	1.6466	-9.6	31.1	-5.44	0
dpropwhite	U	-0.17382	-1.4408	14.2		2.18	0.029
	M	-0.17382	-1.3655	13.3	5.9	7.26	0
dpropbrown	U	-0.72852	-0.1894	-4.8		-0.75	0.453
	M	-0.72852	0.10378	-7.4	-54.4	-3.55	0
dpropblack	U	-0.77739	-1.6363	13.4		2.34	0.019
	M	-0.77739	-1.76	15.4	-14.4	7.62	0
dpropyellow	U	-0.98631	-0.37012	-21.4		-3.31	0.001
	M	-0.98631	-0.96679	-0.7	96.8	-0.36	0.72
dpropindigenous	U	-1.4394	-1.4328	-0.2		-0.03	0.974
	M	-1.4394	-1.1963	-7.2	-3593.9	-4.07	0

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dpropeightyrs_less	U	-0.76388	-1.3848	21.7		3.18	0.002
	M	-0.76388	-0.66527	-3.4	84.1	-2.12	0.034
dpropnineyrs	U	-11.859	-19.827	54.9		9.31	0
	M	-11.859	-13.932	14.3	74	8.06	0
dproptenyrs	U	5.5919	11.312	-43.7		-7.49	0
	M	5.5919	5.8478	-2	95.5	-1.25	0.21
dpropelevenyrs	U	7.8412	8.7959	-8.7		-1.17	0.242
	M	7.8412	9.3635	-13.8	-59.5	-7.5	0
dproptwelveyrs	U	1.2465	2.7547	-27.7		-4.49	0
	M	1.2465	1.0235	4.1	85.2	2.05	0.041
dpropthirteenyrs	U	0.20856	0.6095	-10.9		-1.77	0.076
	M	0.20856	0.36111	-4.1	62	-2.26	0.024
dpropfourteenyrs	U	-0.08314	0.14031	-9.2		-1.56	0.118
	M	-0.08314	-0.09567	0.5	94.4	0.3	0.767
dpropfifteenyrs_more	U	-0.3381	-0.08443	-11.3		-1.78	0.076
	M	-0.3381	-0.26856	-3.1	72.6	-1.77	0.076
dpropworkyes	U	5.9833	16.081	-91.8		-15.42	0
	M	5.9833	7.3125	-12.1	86.8	-6.63	0
dproppresch	U	15.61	32.77	-92.8		-14.26	0
	M	15.61	22.937	-39.6	57.3	-21.96	0
dpropfailed	U	-11.133	-13.376	19.9		3.01	0.003
	M	-11.133	-13.275	19	4.5	10.79	0
dproplivemother	U	3.7864	3.7395	0.8		0.12	0.907
	M	3.7864	4.1609	-6.2	-698.2	-3.27	0.001
dproplivefather	U	3.7138	3.9838	-3.1		-0.5	0.619
	M	3.7138	4.363	-7.5	-140.4	-3.95	0

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dpropnevermothersch	U	-7.4029	-6.5145	-15		-2.43	0.015
	M	-7.4029	-9.7375	39.4	-162.8	18.08	0
dpropnoneight	U	-13.119	-9.863	-44		-6.77	0
	M	-13.119	-14.408	17.4	60.4	9.16	0
dpropnonhighschool	U	1.8287	-0.67503	29.3		4.7	0
	M	1.8287	3.7125	-22	24.8	-12.36	0
dpropparents_tell_study	U	7.5907	8.8665	-14.5		-2.1	0.036
	M	7.5907	6.8245	8.7	39.9	5.07	0
dpropparents_homework_incentive	U	3.724	2.9012	14.1		2.03	0.043
	M	3.724	3.6667	1	93	0.52	0.601
dpropparents_read_incentive	U	4.8825	3.1708	27		3.99	0
	M	4.8825	4.0264	13.5	50	8.03	0
dpropattendclass_incentive	U	5.4107	4.9936	6.9		1.03	0.305
	M	5.4107	5.5581	-2.5	64.7	-1.37	0.172
sch_building	U	1.0499	1.0655	-5		-0.83	0.407
	M	1.0499	1.0614	-3.7	26.4	-2.19	0.028
filtered_water	U	1.4685	1.3745	16.5		2.83	0.005
	M	1.4685	1.4678	0.1	99.3	0.07	0.945
water_provision	U	0.96829	0.93091	17.1		3.38	0.001
	M	0.96829	0.89554	33.3	-94.6	16.06	0
sewer_public	U	0.80624	0.67273	30.7		5.43	0
	M	0.80624	0.64436	37.3	-21.2	20.29	0
sewer_fossa	U	0.22762	0.36727	-30.9		-5.36	0
	M	0.22762	0.41338	-41.1	-33	-22.35	0
garbage_collection	U	0.98563	0.98545	0.1		0.02	0.981
	M	0.98563	0.98765	-1.7	-1053.4	-0.97	0.334

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
principal_room	U	0.95821	0.95636	0.9		0.15	0.881
	M	0.95821	0.97618	-8.9	-873.8	-5.56	0
teachers_room	U	0.93921	0.94182	-1.1		-0.18	0.86
	M	0.93921	0.95487	-6.6	-501.2	-3.85	0
computer_lab	U	0.89924	0.94909	-18.9		-2.71	0.007
	M	0.89924	0.91566	-6.2	67.1	-3.12	0.002
science_lab	U	0.16881	0.36364	-45.1		-8.32	0
	M	0.16881	0.11341	12.8	71.6	8.78	0
special_care_room	U	0.17906	0.25818	-19.2		-3.32	0.001
	M	0.17906	0.14673	7.8	59.2	4.82	0
sports_pitch	U	0.75752	0.74909	2		0.32	0.75
	M	0.75752	0.6697	20.4	-942.3	10.74	0
kitchen	U	0.97704	0.99636	-16.9		-2.13	0.033
	M	0.97704	0.97332	3.3	80.7	1.32	0.188
library	U	0.67228	0.75636	-18.7		-2.92	0.004
	M	0.67228	0.64411	6.3	66.5	3.27	0.001
park	U	0.055	0.02545	15.1		2.13	0.034
	M	0.055	0.0374	9	40.4	4.62	0
dbathroom_outdoor	U	0.01536	-0.00727	7.6		1.56	0.118
	M	0.01536	0.05141	-12.1	-59.3	-7.12	0
bathroom_specialneeds	U	0.26924	0.32364	-11.9		-1.98	0.047
	M	0.26924	0.25683	2.7	77.2	1.55	0.121
special_needs_room	U	0.19739	0.26909	-17		-2.91	0.004
	M	0.19739	0.20124	-0.9	94.6	-0.53	0.596
dn_rooms	U	0.01652	-0.52364	16.5		2.36	0.018
	M	0.01652	-0.11733	4.1	75.2	2.28	0.023

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
tv	U	0.99686	0.98909	9.3		2.14	0.032
	M	0.99686	0.9868	12	-29.4	6.16	0
videocassete	U	0.8373	0.82545	3.2		0.52	0.603
	M	0.8373	0.80972	7.4	-132.9	3.98	0
dvd	U	0.97225	0.97818	-3.8		-0.59	0.556
	M	0.97225	0.98189	-6.2	-62.5	-3.54	0
parabolica	U	0.69343	0.67636	3.7		0.6	0.549
	M	0.69343	0.62402	14.9	-306.8	8.07	0
xerox	U	0.5261	0.48	9.2		1.5	0.134
	M	0.5261	0.48974	7.3	21.1	4	0
retroproj	U	0.8411	0.89818	-17		-2.55	0.011
	M	0.8411	0.90199	-18.1	-6.7	-10.05	0
printer	U	0.9848	0.97455	7.3		1.34	0.18
	M	0.9848	0.98905	-3	58.6	-2.06	0.04
dn_pcs	U	7.4836	4.9091	13.4		1.65	0.098
	M	7.4836	6.7855	3.6	72.9	2.04	0.041
dadm_pcs	U	1.6929	1.1273	17.3		2.64	0.008
	M	1.6929	1.8646	-5.3	69.7	-3.12	0.002
dstud_pcs	U	5.1027	3.9491	6.3		0.77	0.439
	M	5.1027	4.0782	5.6	11.2	3.15	0.002
internet	U	0.92666	0.87273	18		3.31	0.001
	M	0.92666	0.92329	1.1	93.8	0.7	0.482
dn_staff	U	1.3168	0.05818	7.7		1.22	0.223
	M	1.3168	1.521	-1.2	83.8	-0.64	0.519
sch_meal	U	0.98959	0.98182	6.5		1.22	0.221

Continued on next page

Table A6 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	0.98959	0.99032	-0.6	90.6	-0.4	0.687

Notes: Reported are the means of variables for treated and control schools for unmatched and matched sample together with the bias and standardized differences (% bias reduction) and t-tests in matched sample compared to those in unmatched sample. Pseudo r2 before matching was 0.256; after matching, 0.172. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A3: All the PSM techniques used

	One to one			**Nearest neighbour**			**Caliper**			**Kernel**		
Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese												
ATT	7.12***	4.46**	0.23	8.56***	6.09***	11.02*	9.96***	4.36	0.53	6.70***	4.79***	10.60**
	[2.46]	[2.15]	[6.15]	[3.23]	[2.53]	[5.90]	[2.91]	[2.95]	[9.45]	[1.82]	[1.60]	[4.83]
N(control on support)	306	198	14	306	198	14	306	198	14	306	198	14
N(treat on support)	13,009	8,498	1,971	16,803	9,978	23,487	14,822	9,962	23,467	16,803	9,978	23,487
N(total)	17,109	10,176	23,501	17,109	10,176	23,501	17,109	10,160	23,501	17,109	10,176	23,501
(b) Maths												
ATT	4.83*	2.86	5.79*	8.57***	2.63	12.82***	9.25***	2.53	10.24**	3.01	2.41*	10.35***
	[2.90]	[1.81]	[3.29]	[4.26]	[2.12]	[3.17]	[3.92]	[2.48]	[4.90]	[2.40]	[1.34]	[2.60]
N(control on support)	306	198	14	306	198	14	306	198	14	306	198	14
N(treat on support)	13,006	8,498	1,982	16,799	9,978	23,494	14,816	9,962	23,474	16,799	9,978	23,494
N(total)	17,105	10,176	23,494	17,105	10,176	23,508	17,105	10,160	23,508	17,105	10,176	23,508
(c) Dropout												
ATT	1.64	0.76	-0.51	1.34	0.03	0.5	2.14	1.45	-0.42	0.55	1.46*	2.61**
	[1.88]	[1.31]	[2.86]	[2.48]	[1.50]	[2.26]	[2.32]	[1.44]	[2.42]	[1.13]	[0.88]	[1.22]
N(control on support)	308	76	16	308	76	16	308	76	16	308	76	16
N(treat on support)	15,660	4,587	8,733	22,147	5,922	28,160	18,933	5,906	28,127	22,147	5,922	28,160
N(total)	22,455	5,922	28,176	22,455	5,998	28,176	22,455	5,998	28,160	22,455	5,998	28,176
(d) Pass												
ATT	-5.73	-0.05	1.21	-2.11	2.43	0.15	-3.04	2.8	2.66	-0.55	1.02	-0.41
	[2.59]**	[2.84]	[8.12]	[3.00]	[2.62]	[6.23]	[2.75]	[2.82]	[6.65]	[1.36]	[1.48]	[3.37]
N(control on support)	308	76	16	308	74	16	308	74	16	308	74	16
N(treat on support)	15,692	4,587	8,760	22,131	5,910	28,153	18,922	5,883	28,120	22,131	5,910	28,153
N(total)	22,415	5,998	28,119	22,439	5,984	28,169	22,439	5,984	28,136	22,439	5,984	28,169

Notes: Estimations using bootstrap with 500 repetitions. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A4: Propensity Score Matching - Difference in Differences (only states schools)

	One to one			**Nearest Neighbour**			**Caliper**			**Kernel**		
Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese												
ATT	5.2***	6.36**	7.88	4.66**	7.77	12.61*	4.59***	3.16	-4.99	5.70***	6.82**	2.26
	[1.97]	[3.16]	[6.51]	[2.40]	[4.76]	[7.36]	[2.05]	[5.19]	[14.11]	[1.67]	[3.32]	[9.03]
N(control on support)	275	51	8	275	51	8	275	51	8	275	51	8
N(treat on support)	5,105	1,468	190	6,054	3,497	7,688	5,255	3,479	7,464	6,054	3,497	7,688
N(total)	6,329	3,548	7,696	6,329	3,548	7,696	6,329	3,548	7,696	6,329	3,548	7,696
(b) Maths												
ATT	1.32	4.84	-2.13	1.25	7.39	-2.15	1.10	3.65	-10.61	2.03	2.39	-7.08
	[2.52]	[3.42]	[5.33]	[3.19]	[5.07]	[5.21]	[2.69]	[5.81]	[10.35]	[2.21]	[3.53]	[6.40]
N(control on support)	275	51	8	275	51	8	275	51	8	275	51	8
N(treat on support)	5,107	1,468	190	6,054	3,497	7,686	5,257	3,497	7,644	6,054	3,497	7,686
N(total)	6,054	3,548	7,686	6,329	3,548	7,694	6,329	3,548	7,694	6,329	3,548	7,694
(c) Dropout												
ATT	-0.11	-0.79	3.66	-0.57	0.03	5.77**	-0.32	0.66	-0.14	-0.29	-0.13	1.99
	[1.49]	[0.86]	[2.37]	[1.90]	[1.50]	[2.55]	[1.38]	[1.57]	[4.28]	[1.12]	[1.13]	[3.51]
N(control on support)	277	31	8	277	76	8	277	31	8	277	31	8
N(treat on support)	7,100	621	206	8,565	5,922	9,838	7,136	2,216	9,764	8,565	2,238	9,838
N(total)	8,842	2,269	9,846	8,842	5,998	9,846	8,842	2,269	9,846	8,842	2,269	9,846
(d) Pass												
ATT	0.69	-0.66	-1.16	0.56	1.46	-5.64	1.90	2.46	-1.74	1.98	2.04	-3.06
	[1.99]	[2.59]	[7.39]	[2.23]	[3.05]	[8.07]	[1.51]	[4.06]	[13.24]	[1.32]	[2.61]	[10.11]
N(control on support)	277	31	8	277	31	8	277	31	8	277	31	8
N(treat on support)	7,069	621	206	8,557	2,255	9,835	7,133	2,218	9,791	8,557	2,255	9,835
N(total)	8,808	2,269	9,843	8,834	2,286	9,843	8,834	2,286	9,843	8,834	2,286	9,843

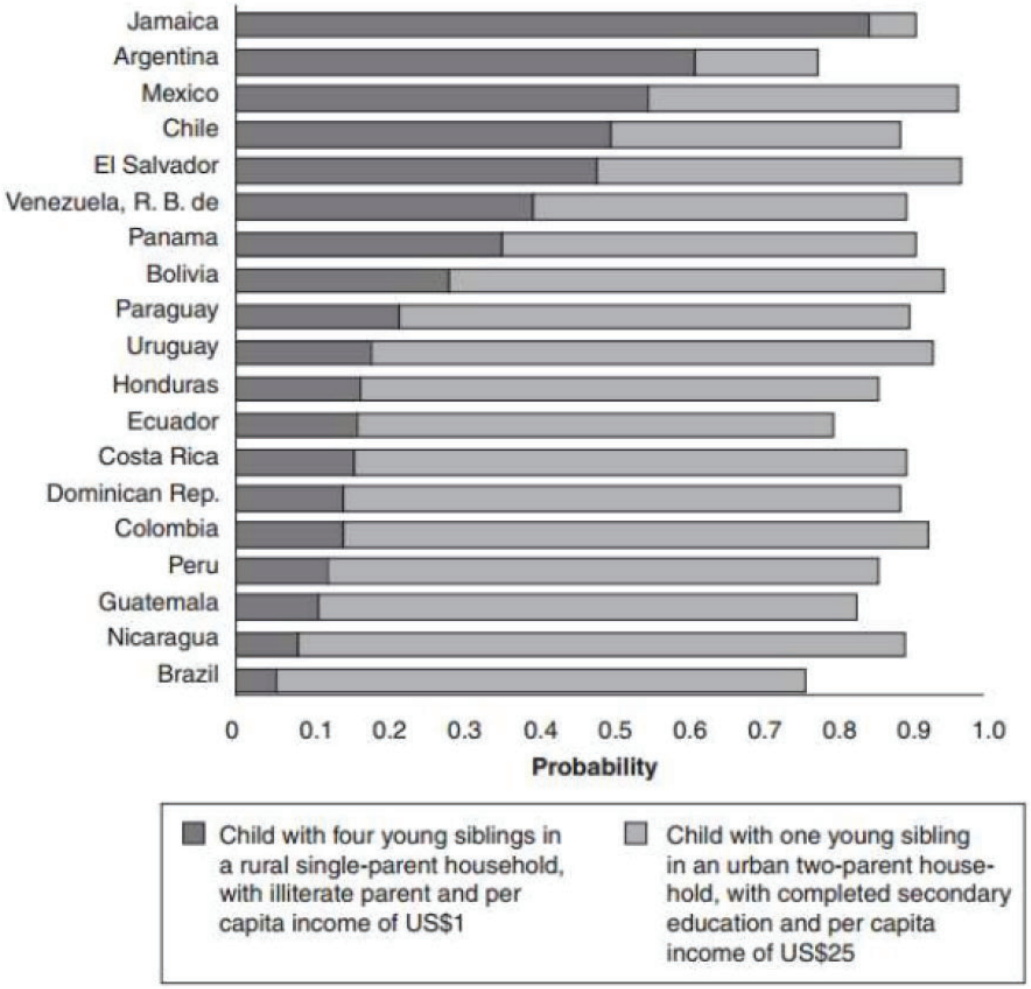
Notes: Estimations using bootstrap with 500 repetitions. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A5: Propensity Score Matching - Difference in Differences (only municipal schools)

	One to one			**Nearest Neighbour**			**Caliper**			**Kernel**		
Dep.Var.	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2	Cohort 1		Cohort 2
	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0	S=0	S=1	S=0
(a) Portuguese												
ATT	4.64	2.83	5.33	2.63	5.46*	10.01	3.8	8.09***	-0.29	3.43*	3.60*	11.86**
	[4.24]	[2.54]	[4	[4.20]	[3.05]	[7.57]	[4.27]	[3.40]	[13.05]	[1.96]	[2.03]	[6.09]
N(control on support)	155	147	14,780	155	147	9	155	147	9	155	147	9
N(treat on support)	24,167	5,363	7	26,921	6,449	18,363	26,896	6,419	18,315	26,921	6,449	18,363
N(total)	27,076	6,596	155	27,076	6,596	18,372	27,076	6,596	18,372	27,076	6,596	18,372
(b) Maths												
ATT	5.32	1.47	13.37***	2.53	1.92	13.05***	3.82	2.45	10	0.84	1.06	12.78***
	[4.75]	[2.02]	[4.18]	[4.70]	[2.43]	[5.13]	[4.79]	[2.72]	[8.90]	[2.21]	[1.62]	[4.14]
N(control on support)	155	147	7	155	147	9	155	147	9	155	147	9
N(treat on support)	24,158	5,363	155	26,913	6,449	18,374	26,889	6,419	18,326	26,913	6,449	18,374
N(total)	27,068	6,449	14,951	27,068	6,596	18,383	27,068	6,596	18,383	27,068	6,596	18,383
(c) Dropout												
ATT	-0.58	0.31	2.64*	-0.65	0.03	3.17*	-0.50	9.98***	3.48	1.29	3.96***	1.28
	[2.06]	[1.39]	[1.56]	[2.21]	[1.50]	[1.83]	[2.11]	[3.98]	[3.10]	[0.94]	[1.68]	[1.16]
N(control on support)	157	45	9	157	76	11	157	45	11	157	45	11
N(treat on support)	28,328	1,112	557	32,920	5,998	20,984	32,892	3,446	20,964	32,920	3,463	20,984
N(total)	33,077	3,508	17,364	33,077	5,998	20,995	33,077	3,508	20,995	33,077	3,508	20,995
(d) Pass												
ATT	1.23	2.46	-8.29	-4.02	1.68	-2.26	-3.97	5.39	-8.02	-4.58	1.20	1.1
	[3.80]	[2.05]	[3.41]***	[3.17]	[4.31]	[7.36]	[3.13]	[5.92]	[13.08]	[1.37]***	[2.79]	[4.65]
N(control on support)	157	43	9	157	43	11	157	43	11	157	43	11
N(treat on support)	28,303	1,097	557	32,882	3,434	20,975	32,840	3,406	20,995	32,882	3,434	20,975
N(total)	33,049	3,477	17,348	33,039	3,477	20,986	33,039	3,477	20,986	33,039	3,477	20,986

Notes: Estimations using bootstrapping to calculate standard errors (SE) with 500 repetitions. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Figure A1: Simulated probability of completion sixth grade on time, *circa* 2005



Source: (Barros, 2015). Note: Brazil legs behind most of its Latin American neighbours when it comes to schooling completion and dropping out probabilities.

Table A7: Matching diagnostic - treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; whole sample)

	Unmatched	Mean			Mean	%reduct	t-test	$p > \|t\|$
Variable	Matched	Treated	Control	%bias	bias			
white	U	-6.0955	-3.3689	-23.70			-2.99	0.003
	M	-6.0955	-4.88	-10.50		55.4	-7.17	0
brown	U	0.50467	-2.5569	20.50			2.75	0.006

Continued on next page

Table A7 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	0.50467	-1.0806	10.60	48.2	7.76	0
black	U	-1.2365	-1.2567	0.30		0.04	0.967
	M	-1.2365	-0.7404	-7.60	-2358	-5.27	0
asian	U	-1.0448	-1.3074	7.9	7.90	1.06	0.29
	M	-1.0448	-0.92288	-3.7	53.6	-2.64	0.008
indigenous	U	-0.92646	-1.2471	9.2		1.22	0.223
	M	-0.92646	-1.1159	5.4	40.9	4.09	0
8 or less years old	U	-1.9752	-2.0168	1.3		0.17	0.865
	M	-1.9752	-2.0692	3	-125.7	2.13	0.033
9 years old	U	-20.924	-17.48	-32.8		-4.46	0
	M	-20.924	-18.674	-21.4	34.7	-14.68	0
10 years old	U	-3.4681	-1.758	-11.4		-1.61	0.107
	M	-3.4681	-3.1116	-2.4	79.1	-1.64	0.101
11 years old	U	26.315	23.715	17.8		2.31	0.021
	M	26.315	25.728	4	77.4	2.75	0.006
12 years old	U	2.2103	1.9681	3.4		0.48	0.632
	M	2.2103	1.1062	15.6	-356	10.95	0
13 years old	U	0.52037	0.07584	8.4		1.31	0.191
	M	0.52037	0.3824	2.6	69.2	1.89	0.058
14 years old	U	0.20957	-0.7809	31.1		4.28	0
	M	0.20957	0.00775	6.3	79.6	4.64	0
15 or more years old	U	-0.60228	-1.0389	11.3		1.66	0.096
	M	-0.60228	-1.1512	14.3	-25.7	10.3	0
worker	U	-3.8841	-3.0412	-8.5		-1.23	0.01
	M	-3.8841	-4.3495	4.7	44.8	3.56	0
attended preschool	U	-2.7417	0.61704	-19.1		-2.58	0.01

Continued on next page

Table A7 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	-2.7417	-0.4522	-13	31.8	-9.11	0
ever failed	U	-7.0169	-6.4208	-4.6		-0.64	0.524
	M	-7.0169	-7.8083	6.2	-32.8	4.45	0
live with mother	U	5.7703	3.9142	16		1.9	0.057
	M	5.7703	4.5399	10.6	33.7	7.45	0
live with father	U	0.50045	-0.49298	8.4		1.06	0.289
	M	0.50045	-0.76266	10.6	-27.1	7.33	0
mother never school	U	-9.001	-10.695	23.8		3.53	0
	M	-9.001	-10.086	15.2	36	11.24	0
mother less 8th grade	U	-13.381	-11.308	-23.3		-3.14	0.002
	M	-13.381	-12.974	-4.6	80.4	-3.3	0.001
mother less high school	U	-1.4628	-0.80645	-7.4		-1.03	0.304
	M	-1.4628	-2.0476	6.6	10.9	4.64	0
parents tell to study	U	6.8404	7.5854	-8.6		-1.16	0.245
	M	6.8404	6.7677	0.8	90.2	0.62	0.537
parents homework incentive	U	5.3361	3.4125	25.2		3.2	0.001
	M	5.3361	4.5839	9.8	60.9	6.94	0
parents read incentive	U	7.4657	5.7551	19		2.3	0.021
	M	7.4657	6.2937	13	31.5	9.24	0
paernts class incentive	U	5.9425	3.0418	32.3		4.01	0
	M	5.9425	4.8555	12.1	62.5	9.34	0
school building	U	0.97284	0.95455	9.8		1.56	0.119
	M	0.97284	0.97499	-1.1	88.2	-0.95	0.341
filtered water	U	0.75476	0.90909	-42.1		-5.02	0
	M	0.75476	0.84373	-24.3	42.4	-15.78	0
kitchen	U	0.98637	0.98485	1.3		0.18	0.855

Continued on next page

Table A7 – continued from previous page

Variable	Unmatched Matched	Mean Treated	Control	%bias	Mean %reduct bias	t-test	$p > \|t\|$
	M	0.98637	0.98941	-2.6	-100	-1.96	0.049
library	U	0.8181	0.73737	19.5		2.91	0.004
	M	0.8181	0.78944	6.9	64.5	5.1	0
park	U	0.3168	0.10101	55		6.5	0
	M	0.3168	0.21148	26.8	51.2	16.99	0
bathroom outdoor	U	0.01824	-0.0202	12.2		1.71	0.088
	M	0.01824	0.03197	-4.3	64.3	-2.81	0.005
bathroom special needs	U	0.47334	0.44444	5.8		0.81	0.42
	M	0.47334	0.47028	0.6	89.4	0.43	0.665
special needs room	U	0.4082	0.36869	8.1		1.12	0.263
	M	0.4082	0.43715	-5.9	26.7	-4.14	0
number of rooms	U	0.48908	0.26263	4.7		0.78	0.436
	M	0.48908	0.39727	1.9	59.4	1.03	0.304
tv	U	0.98878	0.98485	3.4		0.52	0.605
	M	0.98878	0.98582	2.6	24.8	1.86	0.063
videocassete	U	0.6367	0.55556	16.6		2.35	0.019
	M	0.6367	0.61186	5.1	69.4	3.62	0
dvd	U	0.93325	0.89394	14		2.18	0.029
	M	0.93325	0.92841	1.7	87.7	1.35	0.177
parabolic	U	0.39707	0.44444	-9.6		-1.35	0.178
	M	0.39707	0.38931	1.6	83.6	1.12	0.261
xerox	U	0.67769	0.51515	33.5		4.84	0
	M	0.67769	0.65556	4.6	86.4	3.32	0.001
overhead projector	U	0.77931	0.70707	16.6		2.42	0.015
	M	0.77931	0.74457	8	51.9	5.77	0
printer	U	0.94488	0.93434	4.4		0.64	0.521

Continued on next page

Table A7 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	0.94488	0.93986	2.1	52.3	1.52	0.128
number of pcs	U	13.775	12.485	4.5		0.48	0.634
	M	13.775	13.9	-0.4	90.3	-0.31	0.759
staff members	U	4.2504	5.9697	-7.2		-1.12	0.263
	M	4.2504	3.6294	2.6	63.9	1.82	0.069
school meal	U	0.97555	0.95455	11.4		1.88	0.06
	M	0.97555	0.95825	9.4	17.6	6.84	0.000

Notes: Reported are the means of variables for treated and control schools for unmatched and matched sample together with the bias and standardized differences (% bias reduction) and t-tests in matched sample compared to those in unmatched sample. Pseudo R-squared unmatched: 0.083; after matching: 0.060. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A8: Matching diagnostic - treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; only state schools)

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dpropwhite	U	-9.8676	-4.4294	-44.5		-2.95	0.003
	M	-9.8676	-15.88	49.2	-10.6	18.37	0
dpropbrown	U	0.33063	0.29438	0.3		0.02	0.986
	M	0.33063	4.9814	-35.2	12728.5	-13.22	0
dpropblack	U	-1.5229	-1.5119	-0.2		-0.01	0.99
	M	-1.5229	-0.22016	-22.6	11761.8	-8.62	0

Continued on next page

Table A8 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dpropyellow	U	-1.1765	-1.8419	19.8		1.41	0.158
	M	-1.1765	-2.4886	39.1	-97.2	18.93	0
dpropindigenous	U	-0.80761	-1.5868	25.2		1.64	0.101
	M	-0.80761	0.11438	-29.8	-18.3	-13.36	0
dpropeightyrs_less	U	-2.3408	-1.7515	-21		-1.48	0.14
	M	-2.3408	-3.1575	29.1	-38.6	8.81	0
dpropnineyrs	U	-26.301	-23.545	-27.5		-1.84	0.066
	M	-26.301	-19.818	-64.7	-135.3	-28.76	0
dproptenyrs	U	-4.312	-5.0022	4.1		0.32	0.747
	M	-4.312	-5.9572	9.7	-138.4	4.32	0
dpropelevenyrs	U	30.131	29.004	7.3		0.5	0.615
	M	30.131	30.873	-4.8	34.1	-2.33	0.02
dproptwelveyrs	U	2.6934	3.0587	-5.6		-0.38	0.702
	M	2.6934	1.5072	18.3	-224.7	7.86	0
dpropthirteenyrs	U	0.87114	1.152	-6.1		-0.47	0.641
	M	0.87114	-0.06298	20.4	-232.5	10.91	0
dpropfourteenyrs	U	0.44007	-0.5832	38.8		2.59	0.01
	M	0.44007	-1.9526	90.7	-133.8	34.07	0
dpropfifteenyrs_more	U	-0.03611	0.35363	-19.2		-1.13	0.259
	M	-0.03611	-0.1098	3.6	81.1	1.6	0.109
dpropworkyes	U	-4.5475	-1.1249	-36.4		-2.66	0.008
	M	-4.5475	-5.2647	7.6	79	3.35	0.001
dproppresch	U	-6.053	-3.0195	-18		-1.22	0.223
	M	-6.053	1.8636	-47	-161	-21.2	0
dpropfailed	U	-9.0934	-5.1382	-35		-2.33	0.02
	M	-9.0934	-11.75	23.5	32.8	9.49	0

Continued on next page

Table A8 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dproplivemother	U	1.2183	1.7773	-7		-0.58	0.561
	M	1.2183	-0.19432	17.6	-152.7	9.18	0
dproplivefather	U	-2.0537	-2.4981	4.2		0.29	0.769
	M	-2.0537	-8.118	57.7	-1264.7	22.62	0
dpropnevermothersch	U	-7.0631	-6.3135	-14		-0.98	0.329
	M	-7.0631	-10.38	62.1	-342.5	24.53	0
dpropnoneight	U	-13.933	-12.855	-11.3		-0.82	0.411
	M	-13.933	-18.911	52.3	-361.9	20.72	0
dpropnonhighschool	U	-3.4496	-4.8231	17.5		1.15	0.252
	M	-3.4496	-0.32133	-39.9	-127.8	-14.23	0
dpropparents_tell_study	U	3.8795	8.9173	-61.4		-5.67	0
	M	3.8795	3.3114	6.9	88.7	3.71	0
dpropparents_homework_incentive	U	3.6223	5.0511	-25.8		-1.84	0.066
	M	3.6223	4.7604	-20.5	20.3	-10.05	0
dpropparents_read_incentive	U	4.9377	7.0108	-28.7		-2.28	0.023
	M	4.9377	4.3967	7.5	73.9	3.22	0.001
dpropattendclass_incentive	U	3.184	4.5575	-24.4		-1.74	0.081
	M	3.184	3.665	-8.5	65	-4.41	0
sch_building	U	0.97998	0.92157	27		2.9	0.004
	M	0.97998	0.98952	-4.4	83.7	-3.25	0.001
filtered_water	U	0.60967	0.86275	-59.7		-3.69	0
	M	0.60967	0.69057	-19.1	68	-7.12	0
library	U	0.89105	0.84314	14.1		1.09	0.277
	M	0.89105	0.88104	2.9	79.1	1.32	0.188
park	U	0.39291	0.09804	72.7		4.3	0
	M	0.39291	0.28161	27.4	62.3	9.91	0

Continued on next page

Table A8 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
dbathroom_outdoor	U	0.00515	-0.03922	17.1		1.02	0.307
	M	0.00515	-0.0024	2.9	83	1.43	0.154
bathroom_specialneeds	U	0.47984	0.39216	17.7		1.24	0.213
	M	0.47984	0.45057	5.9	66.6	2.45	0.014
special_needs_room	U	0.41607	0.39216	4.9		0.34	0.731
	M	0.41607	0.4114	0.9	80.5	0.4	0.692
dn_rooms	U	0.10237	-0.88235	31.3		2.25	0.024
	M	0.10237	0.26719	-5.2	83.3	-2.64	0.008
tv	U	0.99457	0.96078	22.8		3.13	0.002
	M	0.99457	0.99914	-3.1	86.5	-3.42	0.001
videocassete	U	0.70403	0.62745	16.2		1.19	0.235
	M	0.70403	0.67367	6.4	60.3	2.74	0.006
dvd	U	0.94309	0.96078	-8.2		-0.54	0.588
	M	0.94309	0.9949	-24.1	-192.9	-12.64	0
parabolica	U	0.43752	0.5098	-14.4		-1.03	0.302
	M	0.43752	0.32249	23	-59.1	9.98	0
xerox	U	0.78381	0.56863	47		3.69	0
	M	0.78381	0.91246	-28.1	40.2	-15.23	0
retroproj	U	0.87046	0.88235	-3.6		-0.25	0.802
	M	0.87046	0.9735	-31.2	-766.4	-16.37	0
printer	U	0.97312	0.98039	-4.8		-0.32	0.75
	M	0.97312	0.99644	-15.4	-220.7	-8	0
dn_pcs	U	15.959	14.078	13.8		1.02	0.309
	M	15.959	11.079	35.8	-159.5	17.15	0
dn_staff	U	2.0309	12.549	-36		-3.88	0
	M	2.0309	5.03	-10.3	71.5	-6.19	0

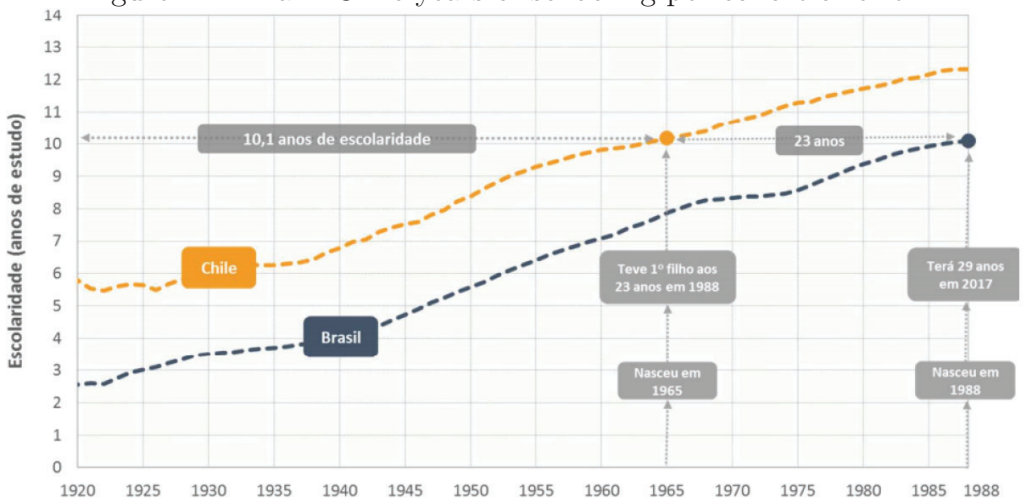
Continued on next page

Table A8 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	Mean %reduct bias	t-test	$p > \|t\|$
sch_meal	U	0.97226	0.96078	6.3		0.49	0.621
	M	0.97226	0.99233	-11.1	-74.9	-6.38	0

Notes: Reported are the means of variables for treated and control schools for unmatched and matched sample together with the bias and standardized differences (% bias reduction) and t-tests in matched sample compared to those in unmatched sample. Pseudo r2 from 0.237 to 0.485 after matching. Levels of significance: * 0.10; ** 0.05; *** 0.01.

Figure A2: Brazil-Chile years of schooling per cohort of birth



Source: (Barros, 2015). Note: The level of schooling (ordinate in the graph) of Brazilian young adults (29 years old) is the same as the Chilean youngsters' parents when they were in the same age (29 years old) 23 years ago. That is, Brazil is lagging one generation behind.

Table A9: Matching diagnostic - Treatment: compulsory school entrance age (Kernel method) cohort 1; s=1; only municipal schools)

	Unmatched	Mean		%reduct		
Variable	Matched	Treated	Control	%bias	bias	t-test $p > \|t\|$
dpropwhite	U	-4.0583	-3.001	-9.6		-1.05 0.293
	M	-4.0583	-4.2476	1.7	82.1	0.94 0.349
dpropbrown	U	0.62304	-3.5462	26.9		3.11 0.002
	M	0.62304	-1.0864	11	59	6.26 0
dpropblack	U	-1.0789	-1.1681	1.3		0.15 0.881
	M	-1.0789	-0.59497	-7	-442.8	-3.96 0
dpropyellow	U	-0.97639	-1.1219	4.4		0.5 0.62
	M	-0.97639	-0.83878	-4.1	5.5	-2.31 0.021
dpropindigenous	U	-0.99325	-1.1293	3.7		0.43 0.669
	M	-0.99325	-1.1935	5.5	-47.2	3.29 0.001
dpropeightyrs_less	U	-1.7712	-2.1089	10.2		1.11 0.269

Continued on next page

Table A9 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	-1.7712	-1.9745	6.2	39.8	3.69	0
dpropnineyrs	U	-17.967	-15.376	-26.9		-3.23	0.001
	M	-17.967	-16.099	-19.4	27.9	-10.5	0
dproptenyrs	U	-3.0201	-0.63327	-16.9		-1.97	0.049
	M	-3.0201	-3.257	1.7	90.1	0.93	0.352
dpropelevenyrs	U	24.223	21.88	16.9		1.85	0.065
	M	24.223	24.041	1.3	92.2	0.7	0.484
dproptwelveyrs	U	1.9451	1.5898	4.9		0.59	0.553
	M	1.9451	1.0032	12.9	-165.1	7.26	0
dpropthirteenyrs	U	0.33005	-0.29754	11.3		1.51	0.13
	M	0.33005	0.42591	-1.7	84.7	-0.97	0.33
dpropfourteenyrs	U	0.08568	-0.8495	27.6		3.27	0.001
	M	0.08568	0.1537	-2	92.7	-1.2	0.232
dpropfifteenyrs_more	U	-0.91598	-1.522	14		1.76	0.079
	M	-0.91598	-1.7105	18.4	-31.1	11.07	0
dpropworkyes	U	-3.5257	-3.7061	1.8		0.22	0.824
	M	-3.5257	-4.5028	9.6	-441.6	5.89	0
dproppresch	U	-0.96542	1.8787	-16.1		-1.89	0.059
	M	-0.96542	1.7517	-15.4	4.5	-8.43	0
dpropfailed	U	-5.8834	-6.8658	7.4		0.88	0.38
	M	-5.8834	-7.4106	11.5	-55.5	6.65	0
dproplivemother	U	8.258	4.6556	28.1		2.77	0.006
	M	8.258	5.5965	20.7	26.1	11.84	0
dproplivefather	U	1.8912	0.20266	13.6		1.45	0.147
	M	1.8912	0.82235	8.6	36.7	4.91	0
dpropnevermothersch	U	-10.069	-12.216	29.2		3.67	0

Continued on next page

Table A9 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	-10.069	-11.328	17.1	41.3	10.18	0
dpropnoneight	U	-13.074	-10.772	-26.6		-3.03	0.002
	M	-13.074	-11.791	-14.8	44.3	-8.73	0
dpropnonhighschool	U	-0.36581	0.58706	-10.6		-1.28	0.199
	M	-0.36581	-1.42	11.8	-10.6	6.89	0
dpropparents_tell_study	U	8.4474	7.1232	15		1.64	0.102
	M	8.4474	7.5317	10.4	30.8	6.08	0
dpropparents_homework_incentive	U	6.2709	2.844	40.9		4.36	0
	M	6.2709	4.7088	18.6	54.4	10.48	0
dpropparents_read_incentive	U	8.8472	5.3194	36.2		3.61	0
	M	8.8472	6.4413	24.7	31.8	14.08	0
dpropattendclass_incentive	U	7.4451	2.5159	49.1		5.13	0
	M	7.4451	5.2303	22.1	55.1	13.71	0
sch_building	U	0.96899	0.96599	1.7		0.21	0.836
	M	0.96899	0.97408	-2.9	-69.8	-1.74	0.082
filtered_water	U	0.83362	0.92517	-28.4		-2.96	0.003
	M	0.83362	0.87765	-13.6	51.9	-7.13	0
kitchen	U	0.98387	0.97959	3.2		0.41	0.685
	M	0.98387	0.98902	-3.8	-20.2	-2.53	0.011
library	U	0.77857	0.70068	17.8		2.24	0.025
	M	0.77857	0.76401	3.3	81.3	1.97	0.049
park	U	0.27508	0.10204	45.3		4.67	0
	M	0.27508	0.21012	17	62.5	8.63	0
dbathroom_outdoor	U	0.02528	-0.01361	11.6		1.47	0.142
	M	0.02528	0.06499	-11.9	-2.1	-5.86	0
bathroom_specialneeds	U	0.47124	0.46259	1.7		0.21	0.835

Continued on next page

Table A9 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test	$p > \|t\|$
	M	0.47124	0.47387	-0.5	69.6	-0.3	0.765
special_needs_room	U	0.40487	0.36054	9.1		1.08	0.279
	M	0.40487	0.43732	-6.7	26.8	-3.73	0
dn_rooms	U	0.6984	0.65986	0.7		0.1	0.918
	M	0.6984	1.0151	-6	-721.7	-2.67	0.008
tv	U	0.98573	0.9932	-7.3		-0.76	0.448
	M	0.98573	0.98509	0.6	91.3	0.31	0.759
videocassete	U	0.59978	0.53061	14		1.69	0.091
	M	0.59978	0.57845	4.3	69.2	2.46	0.014
dvd	U	0.92821	0.87075	19.2		2.65	0.008
	M	0.92821	0.92028	2.6	86.2	1.7	0.089
parabolica	U	0.37556	0.42177	-9.4		-1.14	0.253
	M	0.37556	0.36079	3	68	1.74	0.082
xerox	U	0.62025	0.4966	25.1		3.05	0.002
	M	0.62025	0.63153	-2.3	90.9	-1.32	0.186
retroproj	U	0.72911	0.64626	17.9		2.23	0.026
	M	0.72911	0.71935	2.1	88.2	1.24	0.215
printer	U	0.92929	0.91837	4.1		0.51	0.61
	M	0.92929	0.91396	5.8	-40.3	3.24	0.001
dn_pcs	U	12.579	11.932	1.9		0.17	0.866
	M	12.579	13.367	-2.3	-21.7	-1.31	0.192
dn_staff	U	5.3901	3.6871	7.9		0.95	0.342
	M	5.3901	4.6528	3.4	56.7	1.92	0.055
sch_meal	U	0.97767	0.95238	13.8		2.03	0.043
	M	0.97767	0.96158	8.8	36.4	5.33	0

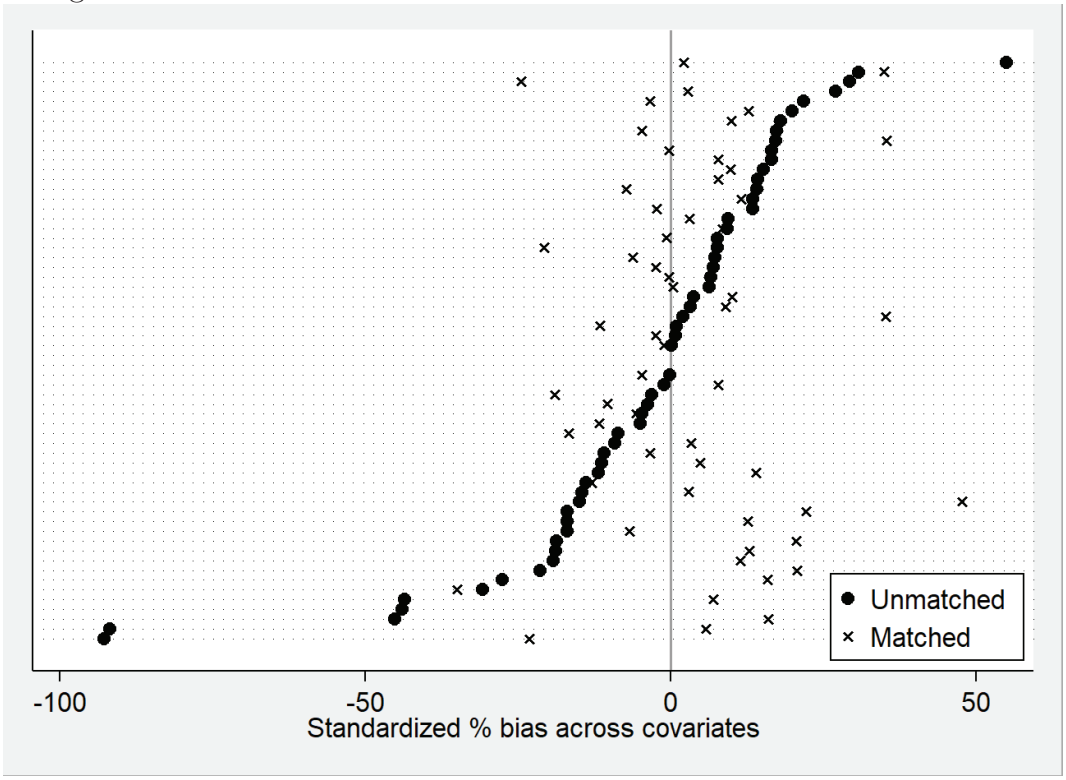
Continued on next page

Table A9 – continued from previous page

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct bias	t-test $p > \|t\|$

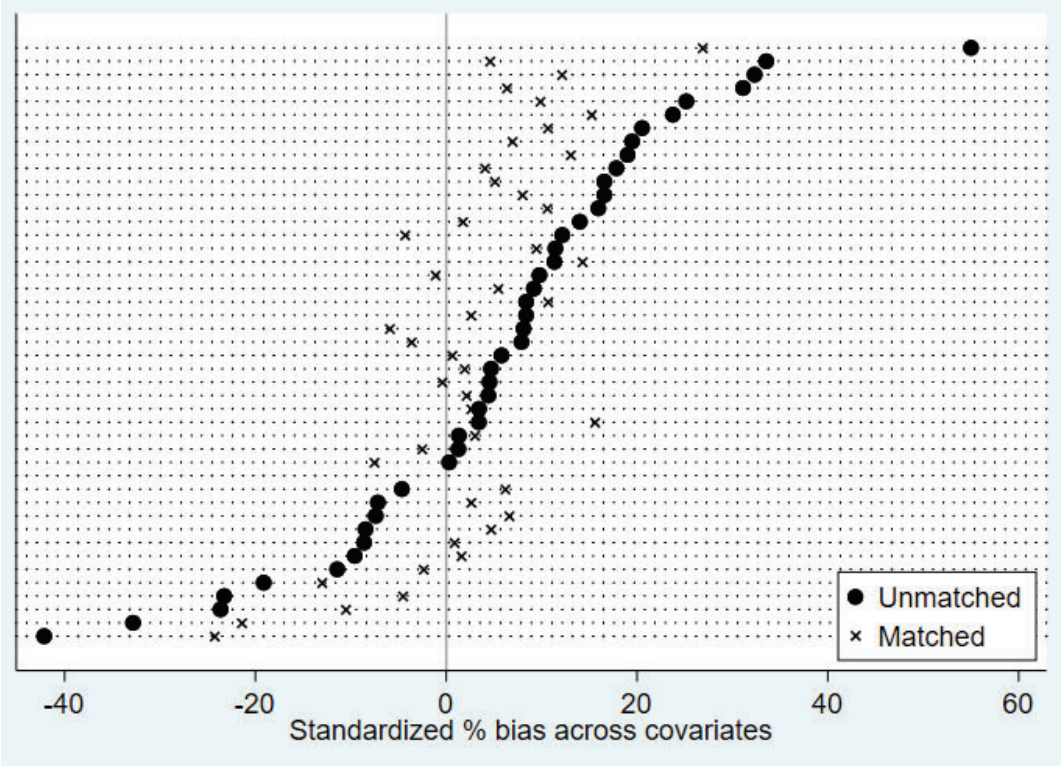
Notes: Reported are the means of variables for treated and control schools for unmatched and matched sample together with the bias and standardized differences (% bias reduction) and t-tests in matched sample compared to those in unmatched sample. Pseudo r2 from 0.088 to 0.080 after matching. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

Figure A3: PSM balancing - state schools only - Method: Kernel - compulsory school entrance age



Notes: The pseudo R-square before the matching was 0.256, and after matching, 0.004. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

Figure A4: PSM balancing - whole sample s=1 - Kernel - compulsory school entrance age



The pseudo R-square before the matching was 0.256, and after matching, 0.004. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

The political-legal education system reforms timeline SEB/MEC (2006):

- Law no. 4,024, of December 20, 1996 - Established 4 years of Elementary School.
- Punta del Este and Santiago Agreement - Commitment to establish six years for Teaching Fundamental until 1970.
- Law no. 5.692, of August 11, 1971 - Obligation of Elementary School of eight years.
- Law 9,394, of December 20, 1996 - admits the enrolment in Primary Education of nine years, the six years of age.
- Law no. 10. 172, of January 9, 2001.
  - Approved the National Education Plan / PNE.
  - A nine-year primary has become the progressive goal of national education.
- Law no. 11. 114, of May 16, 2005 - makes it compulsory for the children of six years of age to age in Elementary School.

Table A10: Whole Sample - Kernel - automatic promotion

	One to one	**NN**	**Caliper**	**Kernel**
Dep. Var.	Cohort 1	Cohort 1	Cohort 1	Cohort 1
	S=0	S=0	S=0	S=0
(a) Portuguese				
ATT	1.4	1.31	0.58	1.29*
	[1.14]	[0.89]	[1.18]	[0.76]
N(control on support)	4,968	4,968	4,968	4,968
N(treat on support)	844	844	836	844
N(total)	5,812	5,812	5,812	5,812
(b) Maths				
ATT	1.54	3.65**	3.73*	4.38***
	[2.37]	[1.92]	[2.28]	[1.57]
N(control on support)	4,968	4,968	4,968	4,968
N(treat on support)	844	844	836	844
N(total)	5,812	5,812	5,804	5,812
(c) Dropout				
ATT	-0.09	-0.14	-0.03	-0.1
	[0.15]	[0.12]	[0.15]	[0.11]
N(control on support)	9,452	9,452	9,452	9,452
N(treat on support)	992	992	986	992
N(total)	10,444	10,444	10,444	10,444
(d) Pass				
ATT	-0.004	0.51	0.19	0.34
	[0.57]	[0.47]	[0.54]	[0.42]
N(control on support)	9,448	9,448	9,448	9,448
N(treat on support)	992	992	985	992
N(total)	10,440	10,440	10,440	10,440

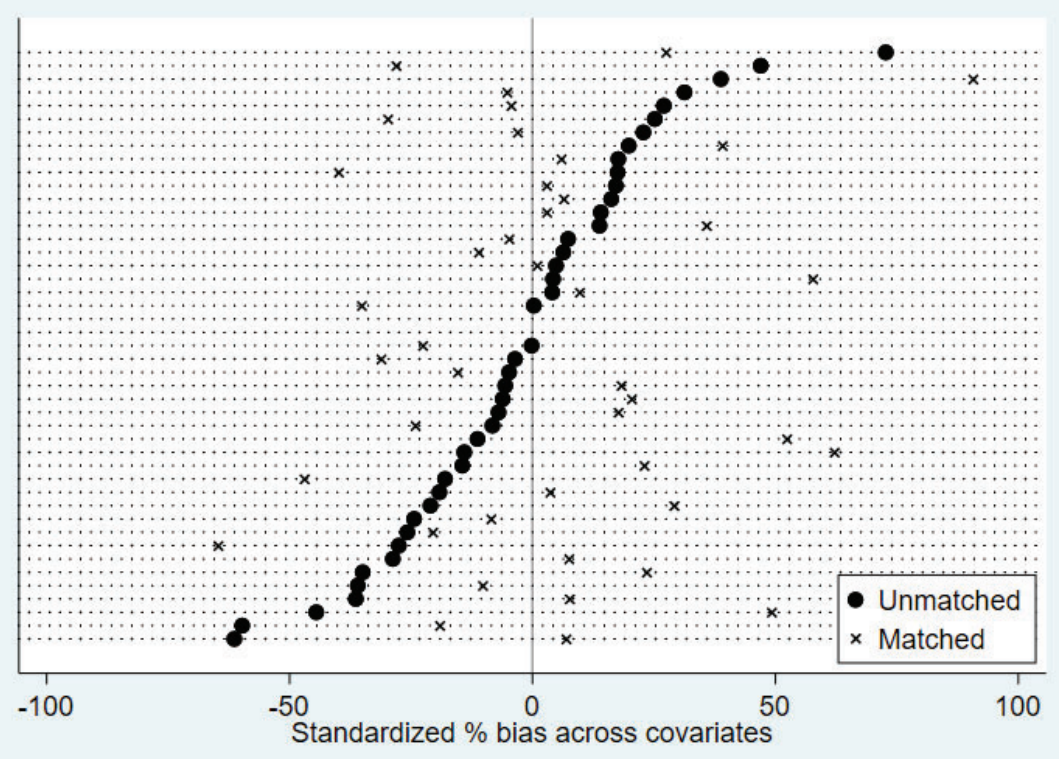
Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works. I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM (I used 500 repetitions). Levels of significance: * 0.10; ** 0.05; *** 0.01.

Table A11: State schools - Kernel - automatic promotion

	One to one	**NN**	**Caliper**	**Kernel**
Dep. Var.	Cohort 1	Cohort 1	Cohort 1	Cohort 1
	S=0	S=0	S=0	S=0
(a) Portuguese				
ATT	3.91	4.22	3.03	1.34
	[5.77]	[10.09]	[7.35]	[9.44]
N(control on support)	765	765	765	765
N(treat on support)	94	140	71	140
N(total)	905	905	905	905
(b) Maths				
ATT	7.92	15.97	-1.72	5.65
	[11.10]	[18.10]	[9.73]	[16.11]
N(control on support)	765	765	765	765
N(treat on support)	94	140	71	140
N(total)	905	905	905	905
(c) Dropout				
ATT	0.75	1.43***	1.18***	1.10***
	[0.48]	[0.51]	[0.48]	[0.30]
N(control on support)	2,951	2,951	2,951	2,951
N(treat on support)	136	144	96	144
N(total)	3,095	3,095	3,095	3,095
(d) Pass				
ATT	3.48***	2.55	2.72	3.80***
	[1.63]	[1.61]	[1.69]	[1.77]
N(control on support)	2,951	2,951	2,951	2,951
N(treat on support)	136	144	96	144
N(total)	3,095	3095	3,095	3,095

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works. I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM (I used 500 repetitions). Levels of significance: * 0.10; ** 0.05; *** 0.01.

Figure A5: PSM balancing -state schools s=1 - Kernel - compulsory school entrance age



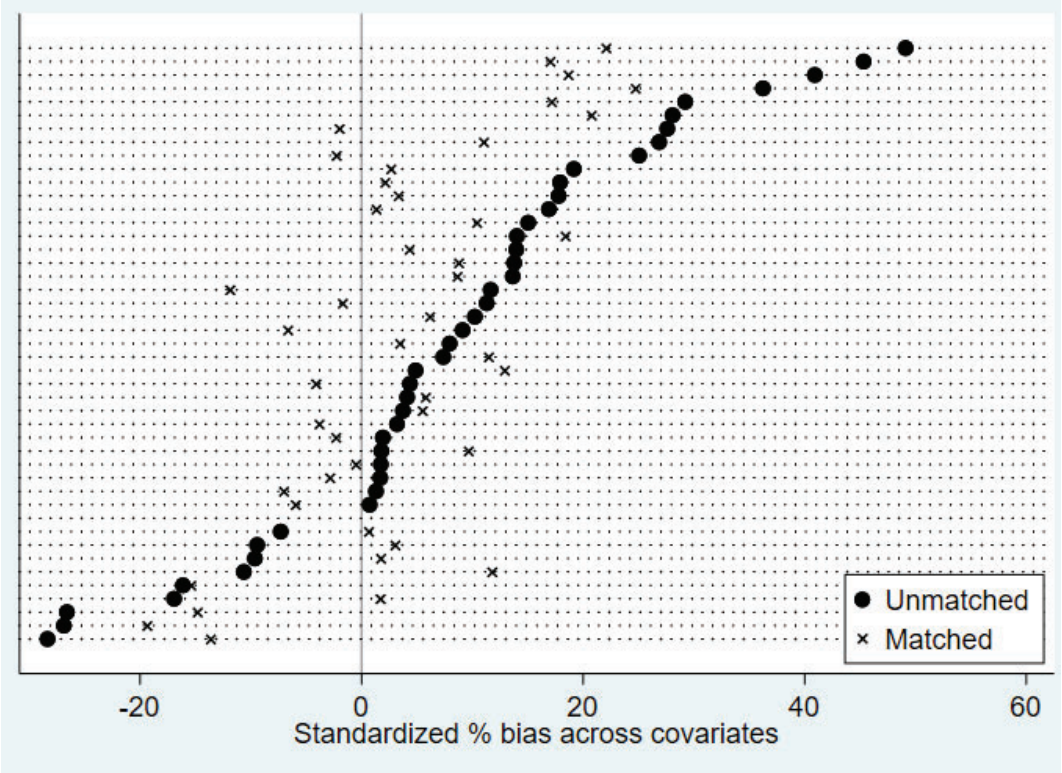
Note: The pseudo R-square before the matching was 0.580, and after matching, 0.223. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

- Law no. 11.274, of February 6, 2006 - extends the Elementary School for a nine-year term, with the enrolment of six-year-old children and establishes a systems by 2010.

Variables for matching and control:

1. **Students characteristics per school**:
    - gender: percentage of students per gender;
    - age: proportion of average age per school;
    - pre-school: percentage of students who attended pre-school;
    - race: proportion of students per race;
    - work: percentage of students who work;
    - percentage of students who ever failed a grade;
    - percentage of students who live with mother/father;

Figure A6: PSM balancing - municipal s=1 - Kernel - compulsory school entrance age



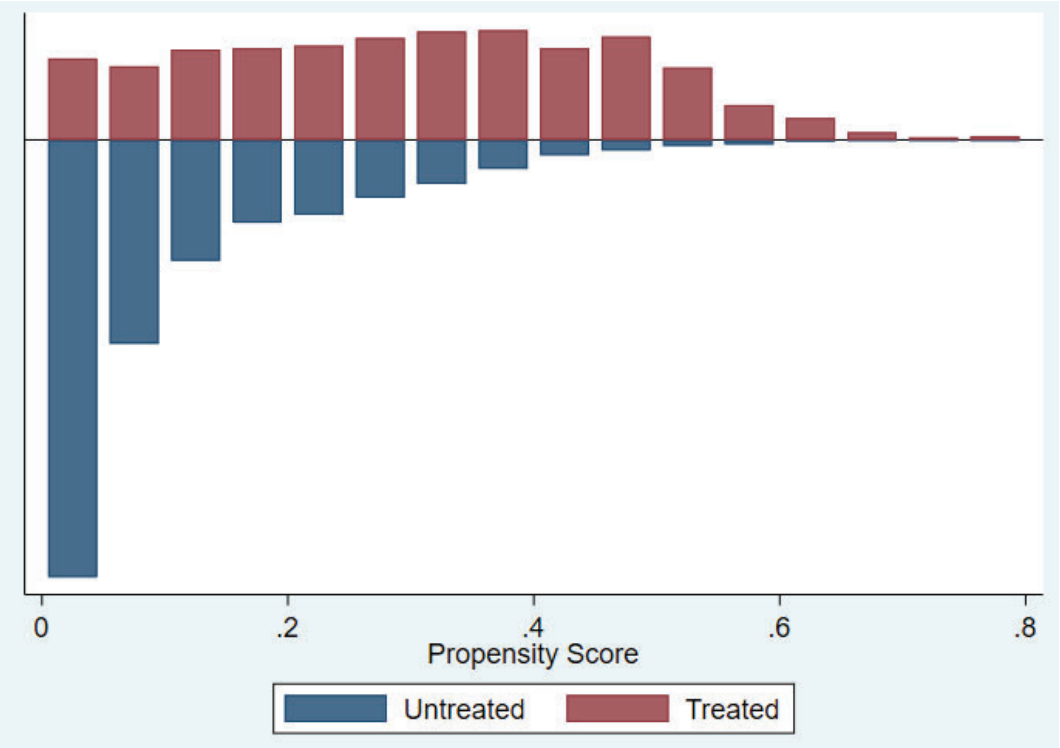
The pseudo R-square before the matching was 0.467, and after matching, 0.214. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

- mother schooling
- parents' interest in children education:
  - parents' incentive to attend the classes;
  - parents' incentive to do the homework;
  - parents' incentive to read;
  - parents' incentive to study.

2. **Schools characteristics**:

- infrastructure:
  - numbers of computers per student;
  - internet access;
  - science laboratory;
  - computer laboratory;
  - teachers room

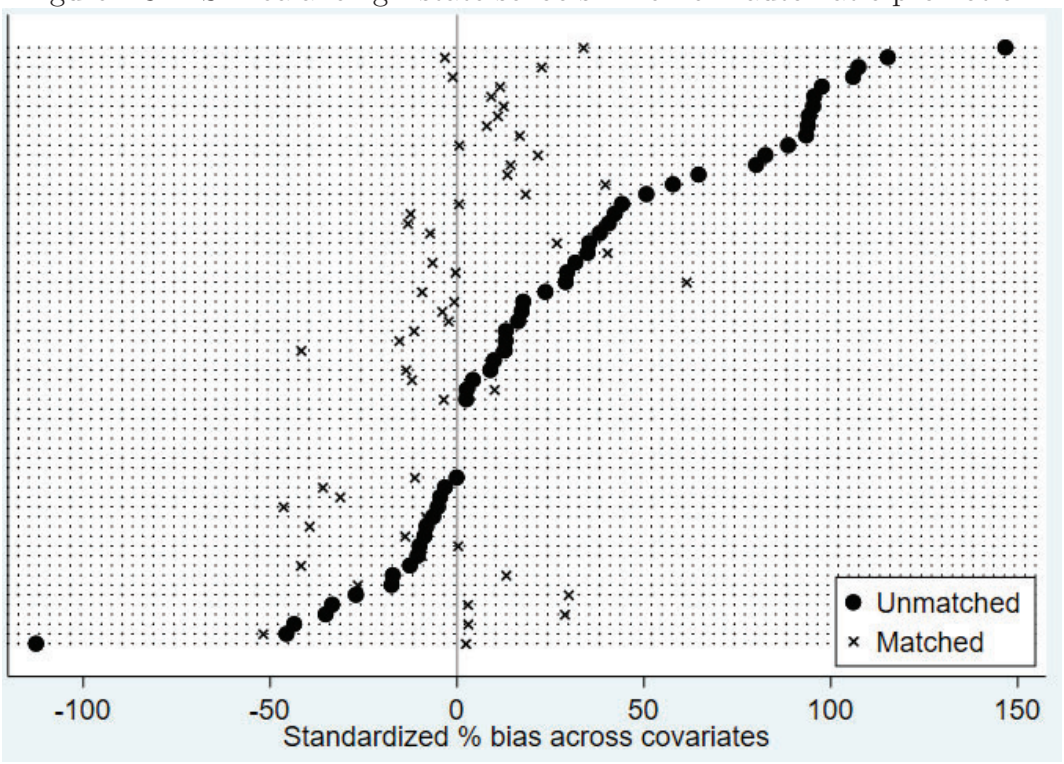
Figure A7: PSM balancing - whole sample - Kernel - automatic promotion)



Notes: The pseudo R-square before the matching was 0.204, and after matching, 0.008. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

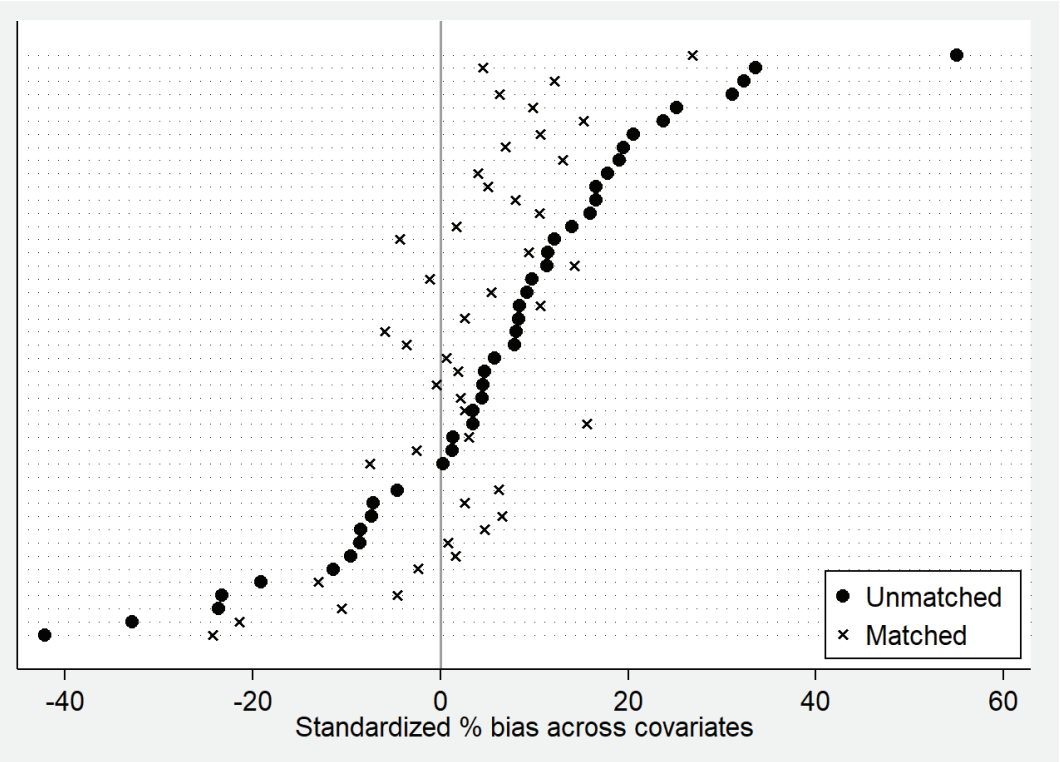
- quality of education:
  - percentage of teachers with university diploma

Figure A8: PSM balancing - state schools - Kernel - automatic promotion



Notes: The pseudo R-square before the matching was 0.629, and after matching, 0.371. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

Figure A9: PSM diagnostic balancing - whole sample s= 1 - Kernel - compulsory school entrance age



The pseudo R-square before the matching was 0.730, and after matching, 0.391. This drop expresses that the confounders inserted into the probit models to estimate the probability of adopting the policy are not different between the control and the intervention groups.

Table A12: Municipal schools - Kernel - automatic promotion

	One to one	**NN**	**Caliper**	**Kernel**
Dep. Var.	Cohort 1	Cohort 1	Cohort 1	Cohort 1
	S=0	S=0	S=0	S=0
(a) Portuguese				
ATT	0.92	0.67	1.33	0.82
	[0.96]	[0.82]	[1.10]	[0.72]
N(control on support)	3,931	3,931	3,931	3,931
N(treat on support)	702	704	701	704
N(total)	4,635	4,635	4,635	4,635
(b) Maths				
ATT	1.3	1.85	2.78	2.61*
	[2.24]	[1.84]	[2.26]	[1.56]
N(control on support)	3,931	3,931	3,931	3,931
N(treat on support)	702	704	701	704
N(total)	4,635	4,635	4,635	4,635
(c) Dropout				
ATT	-0.2	-0.11	-0.12	-0.24
	[0.16]	[0.14]	[0.17]	[0.12]**
N(control on support)	6,128	6,128	6,128	6,128
N(treat on support)	847	848	839	848
N(total)	6,975	6,976	6,976	6,976
(d) Pass				
ATT	-0.2	-0.4	-0.39	-0.21
	[0.62]	[0.51]	[0.59]	[0.46]
N(control on support)	6,124	6,124	6,124	6,124
N(treat on support)	847	848	839	848
N(total)	6,972	6,972	6,972	6,972

Notes: An important detail on PSM estimations is the use of bootstrapping to calculate standard errors (SE) by many empirical works. I follow, for example, Becker et al. (2002) and also use the option bootstrap to calculate the SE of my PSM (I used 500 repetitions). Levels of significance: * 0.10; ** 0.05; *** 0.01.

Chapter 3

The effects of air pollution on students' cognitive performance at Brazilian university entrance tests: Evidence from changes in Wind Direction

3.1 Introduction

The World Health Organisation (WHO) estimates that globally as many as 4.2 million premature deaths each year are linked to exposure to ambient air pollution.[1] The WHO also points out that the health impacts associated with pollution exposure are not equitably distributed, with the poorest individuals in the poorest countries being disproportionately affected. The manner in which many common local air pollutants adversely affect health, particularly via the respiratory and cardiovascular systems, is now well known (Dockery et al., 1993; Hansen and Selte, 2000; Weinhold, 2008; Currie et al., 2009;

[1] http://www9.who.int/airpollution/ambient/health-impacts/en/

Ling and van Eeden, 2009; Currie et al., 2009; Neidell, 2009; Developers, 2016). However, a more recent literature has examined the possibility that exposure to air pollution may also affect cognitive performance and, in turn, potentially human capital formation, productivity, and ultimately economic growth (Zweig et al., 2009; Currie et al., 2014; Ebenstein et al., 2016). If true, this raises the prospect of a 'cycle of disadvantage' in which exposure to pollution constrains the development prospects of those who need it most.

The literature examining the impact of pollution exposure on human capital formation has found compelling evidence that air pollution, through its impact on health, increases school absenteeism (Currie et al., 2009; Ransom and Pope, 2013; Chen et al., 2018; Liu and Salvo, 2018), negatively affects school pupils' scores (Almond and Edlund, 2009; Zweig et al., 2009; Nilsson, 2009; Reyes, 2011; Sanders, 2012; Bharadwaj et al., 2017), and decreases low and high-skilled workers' productivity (Chang et al., 2016; Heyes et al., 2016; Archsmith et al., 2018; Chang et al., 2019; Kahn and Li, 2019). Other studies cast light on the accumulated impact of air pollution within a period of time on pupils' scores at school (Zweig et al., 2009). Overall, these studies look at the long-run impact of fetal (Almond and Edlund, 2009; Sanders, 2012; Bharadwaj et al., 2017) or around birth (Nilsson, 2009; Reyes, 2011)) exposure to air pollution on students' scores and do so typically in developed country settings.

While many studies focus on the cumulative or long-run impacts of air pollution on pupils' academic performance, few pay heed to the contemporaneous effect of pollution exposure on student performance. One exception is Ebenstein et al. (2016) who find negative effects of air pollution exposure on the day of high-stake examinations in Israel, with negative long-run effects on individuals' earnings. Another relevant study is Graff-Zivin et al. (2020) who examine the impact of agricultural fires on Chinese students' exam scores. They differentiate between upwind and downwind fires and find the difference be-

tween the two to have a statistically significant effect on exam performance. With regards to Brazil, Bedi et al. (2020) conduct an experiment at the University of São Paulo and suggest that exposure to high levels of $PM_{2.5}$ reduces students' performance in a reasoning test. Thus, we add to this work by investigating the impacts of particulate matter on students' scores in high-stake exams by using a unique geographic and demographically representative data set. Despite the findings of Graff-Zivin et al. (2020) and Bedi et al. (2020) there remains very limited evidence on the impacts of urban air pollution on students' cognitive performance in the most highly polluted cities and regions of the developing world.

This paper examines the causal relationship between outdoor air pollution levels on the day of the nationwide university entry examinations and students' cognitive performance in Brazil, and specifically in Rio de Janeiro and São Paulo, its most industrialised states. More precisely, we use Brazilian data on concentrations of Ozone (O_3) and particulate matter (PM_{10}) and a rich administrative data set on student's scores to examine the impact of air pollution on academic performance at the national examinations *Exame Nacional do Ensino Médio* (ENEM).[2]. This educational data set is combined with air pollution and weather monitoring station data to build a unique panel data for the period 2015 to 2017. We also test the effects of air pollution on the exam performance of male and female students separately since the previous literature has indicated that fetal exposure to pollution can have different long term impacts on males and females (Jayachandran, 2009; Sanders, 2012). Similarly Ebenstein et al. (2016) found boys' exam performance to be more affected by air pollution than that of girls. In addition, we utilise students' socio-economic characteristics to investigate the heterogenous effects of air pollution on students' scores per different economic strata. In line with other heterogeneity analysis, our results suggest that exam takers from poorer socio-economic backgrounds are more impacted by air pollution than their better-off counterparts.

[2] PM_{10} represents particulate matter with a diameter of 10 micrometers or less.

Our analysis contrasts with Ebenstein et al. (2016) who consider the impact of an overall air quality index on the Israeli national examination composite score, and with Graff-Zivin et al. (2020). The latter work could neither establish the direct link between fires and air pollution nor assess the direct effect of air pollutants on students' scores due to lack of data on air pollution for the period of study. Unlike the previous literature, besides using students' fixed effects, we also implement an instrumental variable approach to mitigate endogeneity concerns related to air pollution. Following Deryugina et al. (2019a) and Bondy et al. (2020) we use wind direction as an exogenous shock capable of producing variation in air pollution levels on a particular day and locality. Our assumption relies on the fact that municipalities' daily wind direction is not directly related to students' scores in the national examinations except through its influence on air pollution. We take advantage of the geographic and temporal variation in wind direction in our data set to explore the exogeneity of our instrumental variable, i.e., the variation in daily wind direction in the municipality where the examination took place.

Our focus on Brazil is motivated by the fact that, by global standards, it experiences relatively high levels of air pollution and relatively low levels of academic performance. In terms of pollution, for instance, in 2015 the Environmental Company of São Paulo (*Companhia Ambiental do Estado de São Paulo* (CETESB)) reported 972 occurrences of ozone levels being above WHO safe limits in the metropolitan area of that state.[3] Regarding academic achievements, in the Organisation for Economic Co-operation and Development (OECD)'s Program for International Student Assessment (PISA) Brazil was ranked 39^{th} out of 40 countries in 2019, revealing the relative underperformance of Brazilian students in comparison with their international counterparts, including most of Brazil's Latin American neighbours. Furthermore, OECD research by Zapata et al. (2015) shows that only approximately 15% of 25-34 year-olds in Brazil completed tertiary education in

[3]The information can be found in the CETESB website: https://cetesb.sp.gov.br/.

2014, compared to the OECD average of 41%.

Our findings suggest that an increase of 10μ/m3 of PM_{10} in the day of the examination decreases students' scores by 6.1 points (8%SD). Male students seem to be more affected than female, with a fall of 8.6 points (11%SD) for boys and of 4.3 points (6%SD) for girls. Furthermore, we find students from poorer households to be more affected by an additional 10μ/m3 of PM_{10}, with scores decreasing by 9%SD for those in the bottom 5% of the income distribution. In contrast, individuals from the top 5% are shown not to be significantly influenced by an increase in pollution levels during the tests. We also provide evidence that weaker students tend to perform even worse than top students, with the former suffering a drop of 6%SD and the latter not significantly affected by air quality deterioration. Finally, we run placebo tests using levels of pollution of one day after the exams, with results that reinforce our main findings that it is indeed the poor air quality in the day of the exams that drives scores to descend. That is, the day after the tests has no significant impact on students' cognitive performance. We therefore believe our study provides policymakers with important evidence of a contemporaneous link between air pollution and exam performance.

The remainder of this chapter is organized as follows. In Section 3.2 we discuss previous literature, while in Section 3.3 we describe our data. We present our empirical strategy in Section 3.4. In Section 3.5, we present our results and conduct sensitivity analysis. In Section 3.6 we present our conclusions.

3.2 Literature on air pollution and cognitive performance

An extensive literature has investigated the impacts of air pollution on human health, human capital formation, productivity and the economic growth of countries. In this

section we summarise the literature on the adverse effects of air pollution exposure on cognitive performance[4]. Researchers have demonstrated that the learning process may be affected by air pollution via the health channel due to its harmful effects on the respiratory system. Although it is known that most pollutants affect the respiratory and cardiovascular systems, as well as damaging the circulatory system by obstructing the flow of oxygen to the brain little is known about the channels through which air pollution harms humans health (Seaton et al., 1995; Currie et al., 2009). PM_{10} small particulates enter in the lung, worsening respiratory and cardiovascular problems such as asthma (Dockery et al., 1993; Hansen and Selte, 2000; Currie et al., 2009; Ling and van Eeden, 2009). Ground-level O_3 can also harm lung function and the respiratory system. Studies have shown that exposure to O_3, and to the pollutants it stems from, is related to premature death, asthma, bronchitis, heart attack, and other cardiopulmonary illnesses (Weinhold, 2008; Currie et al., 2009; Neidell, 2009; Developers, 2016; Wikipedia, 2019). Moreover, other papers have demonstrated that even at low levels of O_3 ($40nmol/mol$), more susceptible individuals are more likely to be harmed when exposed to this pollutant (Bell et al., 2006; Weinhold, 2008). There is also compelling evidence that air pollution, through the health channel, increases school absenteeism (Currie et al., 2009; Ransom and Pope, 2013; Chen et al., 2018; Liu and Salvo, 2018), negatively affects school pupils scores (Zweig et al., 2009; Reyes, 2011; Bharadwaj et al., 2017), and decreases low and high-skilled workers productivity (Chang et al., 2016; Heyes et al., 2016; Archsmith et al., 2018; Chang et al., 2019; Kahn and Li, 2019).

School absenteeism

Currie et al. (2009) examine the effect of air pollution on school absences of elementary and middle school students in Texas. They use a panel data from 1996 to 2001 with schools within 10 miles from the air pollution monitors and differences-in-differences-in-

[4]This literature review is not exhaustive. For a thorough review, please, refer to Currie et al. (2014).

differences to tackle the endogeneity stemming from parents address decisions and find that carbon monoxide (CO) significantly increases absences, but PM_{10} presents mixed effects.

Taking advantage of a natural experiment due to the temporary closure of large local steel mill in Utah Valley in the US, Ransom and Pope (2013) uses the mill closure as an instrument for air pollution and find that that particulate pollution has a strong impact on pupils' absence, while CO does not. Although their Ordinary Least Squares (OLS) results underestimate the effect of particulates on school absenteeism, the IV estimates of the effect of CO are, contrary to expectations, negative effects.

Looking at the impact of air pollution on Chinese students' health and school attendance, Chen et al. (2018) use temperature inversion as an exogenous shock for daily air pollution as an instrumental variable. The study matches the school level data with readings from air pollution monitors within a 40 km radius and find a positive relationship between O_3, $PM_{2.5}$, Sulfur Dioxide (SO_2), CO, and AQI and school absenteeism through the health channel.

In a similar study, Liu and Salvo (2018) build a student level panel to investigate the effects of $PM_{2.5}$ on students' absences from international schools in North China. They find a moderate absence rate stemming from the exposure to $PM_{2.5}$, with United States (US)', Canadian and European students more susceptible to skip school days due to air pollution than their Chinese counterparts. This heterogeneity may be due to the long-run adaptation actions by the Chinese parents who are more used to high levels of pollution in comparison to the aforementioned countries.

School examinations scores and human capital formation

Demonstrating the mechanisms through which air pollution affects students' academic performance is a challenging task as tests scores unveil cumulative knowledge. Therefore researchers need to understand the effects of cumulative exposure to air pollution on students' scores (Currie et al., 2014). One of the mechanisms may be through the impact of air pollution on school absence and, ultimately, academic performance. For example, Zweig et al. (2009) investigate the effects of outdoor air pollution levels on standardised state test scores of California public school pupils using the variation in community pollution over time. Using school fixed effects models they find that a 10% decrease in outdoor PM_{10}, $PM_{2.5}$, or Nitrogen Dioxide (NO_2) would raise Mathematics tests scores by 0.15%, 0.34%, and 0.18%, respectively, while a 10% decrease in outdoor $PM_{2.5}$ increases reading scores by 0.21%.

Using Swedish administrative data, Almond and Edlund (2009) investigate the impact of early life exposure to radioactive particles that came from the Chernobyl nuclear disaster on students' scores. They find that the cohorts who were exposed to radiation in utero had their overall scores reduced by roughly 2.5 percentage points in comparison with the pupils not exposed to the pollutant. Also in Sweden, Nilsson (2009) takes advantage of the leaded gasoline ban during the 1970s to assess the effects of lead exposure around birth on pupils' academic performance. The author finds that the reduction in ambient lead is associated with an increase in high school graduation rates by 2.3%, with higher effects for less well-off students.

Reyes (2011) looks at the long-run effects of lead pollution exposure in early childhood in the 1990s on pupils' tests scores in the 2000s. The author uses individual-level data on childhood lead levels and test scores in Massachusetts, controlling for schools and community characteristics and find that high levels of blood lead in early childhood adversely impact standardised test scores. Sanders (2012) investigates the long-run effects of fetal exposure to Total Suspended Particulates (TSPs) on High School students' scores

in Texas. The author takes advantage of an industrial recession in the early 1980s as an exogenous shock to TSPs to establish the negative causal impacts of this pollutant on the students' scores. He also show evidence that boys present worse academic performance than girls due to the prenatal exposure to TSPs.

Examining the impacts of fetal exposure to air pollution on fourth-grade test scores in Santiago, Chile, Bharadwaj et al. (2017) find a strong negative effect from fetal exposure to CO and correlated pollutants, such as PM_{10}, on Mathematics and language skills. Their identification strategy relies on a plethora of fixed effects to control for parents characteristics and siblings fixed effects, since their data on maternal characteristics misses some crucial unobservable family features that may be related to test outcomes and pollution exposure. Additionally, the work includes controls for air quality alerts to account for time-varying avoidance behaviour the parents may engage in to reduce their exposure to air pollution during periods of higher pollutants emissions. The authors find that the 50% reduction in CO in Santiago between 1990 and 2005 increased lifetime earnings by roughly US$100 million per birth cohort.

Most recently two works have addressed the impacts of contemporaneous air pollution on cognitive performance. Firstly, Ebenstein et al. (2016) investigate the contemporaneous effects of air pollution on students' performance at high stakes examinations in Israel. Controlling for a plethora of fixed effects on students' and parents' characteristics, and weather variables, the authors find that the students receive lower scores when exposed to higher levels of air pollution in the exact day of the exams. Secondly, Bedi et al. (2020) have studied the effects of poor air quality on a series of cognitive test in São Paulo. Their findings suggest that the exposure to high levels of $PM_{2.5}$ causes impaired performance on some of the cognitive abilities tested in their experiment.

Low and high-skilled workers' productivity

As a result of the impacts of air pollution on the learning process through the health channel, there is large literature investigating the effects of air pollution on low and high-skilled workers' productivity. For instance, Chang et al. (2016) look at the impact of outdoor $PM_{2.5}$ on the productivity of indoor pear-packing workers. They find that the pollutant negatively affects worker productivity, whereas O_3 has little or no effect. In the same token, Chang et al. (2019) examine the effect of air pollution on call center workers' productivity in China. Using individual level data and controlling for possible meteorological confounders, the authors find that a 10-unit increase in the Air Pollution Index (API) reduces the number of daily calls (per worker) by 0.35%.

Heyes et al. (2016) investigate the effects of $PM_{2.5}$ on high-skilled workers of the New York Stock Exchange. The authors report that a one standard deviation increases in ambient $PM_{2.5}$ reduces same-day S&P 500 returns by 11.9%, suggesting that air pollution increases the operators' risk aversion via pollution-induced changes in mood or cognitive function. Archsmith et al. (2018) also look at the air pollution impact on work performance of quality-focused baseball umpires. The authors take advantage of the fact that these professionals are quasi-randomly assigned across time and matches venues, to estimate the effects of CO and $PM_{2.5}$ on the umpire's decision making. Using fixed effects and a variety of weather controls, they find that a 1 ppm increase in 3-hour CO increases by 11.5% the likelihood of umpires to make incorrect calls and a 10 μ/m3 increase in 12-hour $PM_{2.5}$ causes a 2.6% increase.

Using individual level data from China, Kahn and Li (2019) show evidence that judges (high-skilled workers) are less productive on more polluted days. After controlling for a variety of fixed effects and weather conditions, the authors find that the duration of lawsuits rises by 19.8% due to an increase of 1% in the indoor level of $PM_{2.5}$ the judges are exposed to. This study also demonstrates that the public sector's productivity is more

impacted by air pollution than the private sector's.

3.3 Data and Summary Statistics

Our data set on daily air quality, weather conditions and students' ENEM scores in São Paulo and Rio de Janeiro between 2015 and 2017 is collected from three main sources as detailed below.

3.3.1 Air pollution data

Air pollution data from São Paulo and Rio de Janeiro is collected from their respective environmental state agencies: the Environmental Company of São Paulo (*Companhia Ambiental do Estado de São Paulo*, CETESB) and the State Institute for the Environment (*Instituto Estadual do Meio Ambiente* (Inea)). Both agencies are responsible for monitoring the air quality of the two states, using ground-level automatic and non-automatic monitoring stations.[5] São Paulo has 63 automatic monitoring stations and 26 non-automatic stations, while Rio de Janeiro has 58 automatic and 116 non-automatic stations (see figure B3). Inea's and CETESB's websites provide the addresses, latitude and longitude of their monitoring stations, which we used to merge this data set with the national examination scores, by municipality centroid location.

Unfortunately, for confidentiality reasons, the Ministry of Education in Brazil does not provide the exact address where each student took their examinations, only the municipality. Hence, we are not able to identify the closest monitoring station to the exact exam location. Nevertheless, following Graff-Zivin et al. (2020) we proxy each exam venue's location using the centroid of the municipality in which the test took place. We then identify up to 3 pollution monitoring stations within 10km of the municipality centroid

[5]The automatic stations are directly linked to a central computer that registers the pollution concentrations. In the non-automatic stations, the samples are manually collected in the site to be analysed in the laboratories of CETESB and Inea.

and average hourly readings across those to form an average daily measure of pollution.[6] We then calculate the Inverted Weighted Distance (IWD) for PM_{10} and O_3 (in microgram per cubic meter, μ/m3) following Currie and Neidell (2005); Bondy et al. (2020) and Sager (2019).

3.3.2 Weather data

To control for atmospheric climate we collected data on weather conditions from CETESB and Inea monitoring stations. This provides us with measures of temperature (in degrees Celsius, ^0C), relative humidity (in percentage, %), wind speed (in meters per second, m/s) and wind direction (in degrees)[7]. The latter is used as an instrumental variable for PM_{10}. CETESB provides wind direction in cardinal mode, thus, to use it as a numerical variable, we transform it into degree direction. Moreover, as wind direction is an angle (360^0 circle) and covaries with wind speed (i.e., there are two components that need to be averaged together) to calculate its daily average we need to split out the East/West vector and the North/South vector, and then recombine them to find the speed and direction (Webmet, 2019; Itron, 2019). [8] The remaining climate measures are also registered on an hourly basis, with the daily means calculated and assigned to the examinations municipalities using the IWD in a similar manner to which the pollution measure is built (Currie and Neidell, 2005; Sager, 2019; Bondy et al., 2020). Table 3.1 presents the summary statistics for weather conditions per exam day.

[6]To be clear, if there are three monitors within 10km we average across those, if there are two we average across those and if there is only one then we use the daily average of that one monitor. If a municipality has no monitoring stations within 10km of its centroid then data for that municipality is dropped from our analysis. In our sensitivity analysis we increase the 10km limit to 50km.

[7]Unfortunately, we do not include rainfall as most stations do not present complete data for this variable for the period analysed in this research.

[8]As in Deryugina et al. (2019a), wind direction refers to the direction the wind is blowing from.

Table 3.1: Summary statistics air pollution, weather and exams scores

Variable	All	Girls	Boys
Exam-level data			
Total Score	526	516	537
(0 - 1000 points)	(76.2)	(71.3)	(78)
Sciences	502	493	515
(0 - 1000 points)	(78.9)	(76.4)	(80.2)
Humanities	558	550	569
(0 - 1000 points)	(77.3)	(76.2)	77.5)
Languages	532	532	534
(0 - 1000 points)	(68.4)	(67.1)	(70)
Mathematics	512	494	535
(0 - 1000 points)	(116)	(106)	(125)
PM_{10} ($\mu g/m^3$)	17.5	17.5	17.5
	(6.11)	(6.13)	(6.08)
O_3 ($\mu g/m^3$)	41.5	41.5	41.6
	(12.3)	(12.3)	(12.1)
Temperature (^0C)	20.6	20.6	20.6
	(3.2)	(3.25)	(3.24)
Humidity (%)	81.4	81.4	81.5
	(13.6)	(13.6)	(13.6)
Wind speed (m/s)	2.25	2.25	2.25
	(1.25)	(1.24)	(1.25)
Observations	2,931,368	1,671,910	1,259,458
N Municipality	36	36	36

Note: Table reports statistics for the estimation sample. Standard deviation between parenthesis. N municipalities attends for the total number of municipalities where the exams took place with weather and air pollution measures available for the three years in our main sample.

3.3.3 University entrance examination scores and other ENEM data

The subject specific ENEM scores are collected from the National Institute for Educational Research *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP), an agency from the Brazilian Ministry of Education (INEP, 2017b). This is a rich administrative data set with detailed information on the students' scores per exam subject at the student and school level, and demographic information, such as household income, parental level of education, age, race, gender and special needs (Table 3.1 summarises this information). We build a repeated cross-sectional dataset at the student level from 2015 to 2017.

The data for 2015 indicates that exams took place in 261 municipalities: 208 municipalities in the state of São Paulo and 53 in Rio de Janeiro state. In 2016, the dataset indicates 209 municipalities in the state of São Paulo and 51 in Rio de Janeiro state, making a total of 260 municipalities where the ENEM took place. In 2017, the state of São Paulo held examinations in 209 municipalities, and the state of Rio de Janeiro in 52. Note that since we use monitoring stations as far as 10km from the municipality centroid to construct our air pollution and meteorological measures, we end up with only 36 municipalities within our final sample. This reduction in the number of municipalities and hence number of spatial observations of pollution measures in our final sample consists of a trade-off decision that we must take. Had we opted for using stations located, for instance, 50km from the municipality centroid, we would have experienced attenuation bias in our estimates [9]. Nonetheless, the reduction in the number of observations in our final panel does not affect the power of our study since we still end up with a very large sample size.

[9]Table B4 in the appendix presents the number of observations and municipalities for the 5 samples using the different distances.

As for demographic information, the INEP data provide the highest level of education achieved by each student's parents. As such, we build dummies indicating each of the seven educational levels. We create eight dummies for each parent's education: (i) Never studied, (ii) Incomplete 1^{st} elementary if the parent attended school for a while, but did not complete the first elementary, (iii) Incomplete 2^{nd} elementary if the parent attended school for a while, but did not complete the second elementary, (iv) Elementary if the parent only completed the Elementary School, (v) High School if the parent completed the High School, (vi) Undergraduate if the parent completed Undergraduate, (vii) Post-Graduate if the parent completed Post-Graduate level, and (viii) non-declared if the students declared they were unaware of their parents' level of education. (Zweig et al., 2009).

Students' socio-economic background is also likely to be an important predictor of their academic performance, particularly in developing countries where the degree of inequality is much higher than in developed nations. In Brazil it has been found that students' socio-economic characteristics can determine up to 85% of the variation in the scores at the exam (Estadao, 2019). In light of that, in our models that use municipality fixed effects, we control for socio-economic background using parental household income. The original data does not display the income in figures, preventing us from using it as a continuous variable. As a result, we created dummies for each level of parental household income, dropping the missing values. According to the INEP's micro-data information, there are 17 income ranges - monthly earnings from zero minimum wage until more than 20 minimum wages -, thus we created 17 dummies, including 16 in our models.

The ENEM

The ENEM is a non-mandatory and standardised national exam used to evaluate high school students in Brazil and is the second largest in the world after the National Higher Education Entrance Examination in China. Since 2009, the exam has been used by fed-

eral and state universities as their admission test. The exam is also used to rank deprived students to receive points in the federal scholarship *Universidade para Todos Program* (or ProUni), as well as for certification for a high school degree.

The ENEM comprises four groups of multiple choice exams that take place in two days: Humanities, Sciences, Languages, and Mathematics, each of them with scores ranging from 0 to 1,000. The Humanities exam includes History, Geography, Philosophy and Sociology, the Sciences tests engulfs Chemistry, Physics and Biology, and the Languages includes Portuguese Language, Literature, Foreign Language (English or Spanish), Arts, Physical Education, Information and Communication Technologies (INEP, 2017b). In 2015 the examinations occurred on the 24^{th} and the 25^{th} October; in 2016, on the 5^{th} and the 6^{th} November; and in 2017, on the 5^{th} and the 12^{th} November. In each year the exams start at 13h30min and finish at 18h30min on the first day, and start at 13h30min and finish at 19h00min on the second day. In 2015 and 2016 on the first day, students took exams on Humanities and Sciences; and in the second day, Languages and Mathematics. In 2017 the order was slightly different, with students taking tests on Languages and Humanities in the first day, and Sciences and Mathematics in the second day. Although we have hourly data for both air pollution and weather variables, which would enable us to assign the measures of air pollution and weather condition at the exact moment of the exams, we follow the literature on the impact of air pollution on human capital and use the daily average of O_3, PM_{10}, humidity, temperature, wind speed, and wind direction (Ebenstein et al., 2016; Sager, 2019; Bondy et al., 2020).

3.4 Empirical Strategy

We examine the partial correlation between PM_{10} exposure and test scores at the individual-level for the years 2015 to 2017. Nonetheless, we know that there are a few

identification challenges to establish the causal relation between air pollution and students' performance due to the possible presence of omitted correlated elements. For instance, one concern is that students' exam scores may be correlated with our two measures of air pollution due to the fact that students from higher socio-economic strata could live in municipalities with lower levels of air pollution (the parents could choose to live in less polluted areas, which are usually more expensive) (Banzhaf and Walsh, 2008; Currie et al., 2014). Hence, although the candidates do not choose the exam venue, it is determined according to the student's household postcode, i.e., INEP tries to assign the examination venue as close as possible (up to 30 km) to each student's address. Hence, students who live in less polluted municipalities would end up taking exams in cities with cleaner air than other candidates who live in more deprived areas, which may be more polluted (Currie et al., 2014). This possibility of address selection is also highlighted by Neidell (2004) and Neidell (2009), who characterise it as an avoidance behaviour by parents.

A final concern relates to the fact that the level of pollution and weather conditions in the days and venues where the exams are marked may influence the final marks as the examiners could also have their productivity affected by these circumstances. However, all the tests are multiple choice and the grades are given by a machine, which clearly is not affected by air pollution or weather determinants. Thus, to tackle those econometric concerns we apply two specification strategies as detailed below.

3.4.1 Panel Fixed Effects Model

As students take exams in two days, we use students fixed effects and control for the exact day of the examinations. The main model thus takes the following form:

$$S_{imt} = \beta PM_{mt} + \tau TEMP_{mt} + \rho HUMID_{mt} + \psi WINDSPEEDmt + \phi_i + \theta_t + \epsilon_{imt} \tag{3.1}$$

where S_{imt} is the test score of student i at municipality m at time t, PM_{mt} is our measure of air pollution (PM_{10}) (in μ/m3) at municipality m at time t, $TEMP_{mt}$ is the daily average temperature at municipality m at time t in degrees Celsius, $HUMID_{mt}$ is the relative humidity measure at municipality m at time t in percentage[10], $WINDSPEED_{mt}$ is the mean of wind speed measured at municipality m at time t in m/s; ϕ_i is students fixed effects. By including fixed effects and weather controls, we are able to compare students' performance in the exams in the different days with different levels of air pollution given each student's characteristics. Hence, confounding factors stemming from time-invariant differences between students can be eliminated from our list of identification concerns. For example, socio-economic background and parents' level of education are subsumed by the students fixed effects. θ_t controls for day of exam fixed effects. As we know the day of each examination, we are able to add this control and rule out the concern that the tiredness of students in the second day of exams would confound the real effects of air pollution on their performance due to extra exhaustion in the second day of tests. In addition, we can eliminate another concern regarding differences in level of pollution from Saturdays and Sundays (when the exams take place) by including these time-varying controls. Finally, the standard errors are clustered by municipality centroid to control for spatial and serial correlation within each municipality in which examinations took place. and ϵ_{imt} is an idiosyncratic error term.

[10]In our Econometric models, we add linear and quadratic terms in both temperature and relative humidity, and their linear and quadratic interaction.

3.4.2 The instrumental variable approach

After using students fixed effects and controlling for weather conditions, there may still remain concerns on the existence of classical measurement error. For example, the students' actual exposure to outdoor air pollution may differ from the monitoring stations readings, which may bias our estimates downward due to attenuation bias (Moretti and Neidell, 2011; Schlenker and Walker, 2015; Chen et al., 2018). Furthermore, as explained above, as PM_{10} is not randomly assigned, leading to the potential presence of unobserved time-varying effects that we cannot account for. Therefore, we also experiment with an instrumental variable approach to estimate our models. In this regard we follow previous works, as, for example, Anderson (2015); Deryugina et al. (2019a); Bondy et al. (2020), and adopt wind direction as our instrument for air pollution, which can be considered as a natural experiment that results in exogenous shocks to local air pollution in a specific area due to wind shocks (Currie et al., 2014).

To identify the causal relationship between pollution and the students' scores we follow the strategy employed by Deryugina et al. (2019a) and Bondy et al. (2020) to build our instrumental variable. To this end we rely on the assumption that municipalities' daily wind direction is not directly related to the students' scores in the national examinations, except through its influence on air pollution. Therefore, as argued by Deryugina et al. (2019a), we do not need to know the exact source of pollution since using wind direction - which is totally random - to predict air quality eliminates the concern of potential correlation between what generates pollution and other factors affecting student performance. We take advantage of the geographic and temporal variation in wind direction in our data set as the source of exogeneity of our instrumental variable. That is, we use the variation in daily wind direction in the municipality where the examination took place as

an instrument for pollution. Our model can be formally described as:

$$Pollutant_{imt} = \sum_{g \in G} \sum_{b=0}^{2} \beta_b^g 1\left[G_c = g\right] \times WindDir_{mt}^{90b} + \theta_1 temp_{mt} + \theta_2 humidity_{mt} + \theta_3 windspeed_{mt} + \nu_i + \phi_t + \epsilon_{imt} \tag{3.2}$$

$$Score_{imt} = \alpha \widehat{Pollutant}_{mt} + \psi_1 temp_{mt} + \psi_2 humidity_{mt} + \psi_3 windspeed_{mt} + \mu_i + \omega_t + \upsilon_{imt} \tag{3.3}$$

where $Score_{imt}$ is the test score of student i at municipality m at time t. The variable $1\left[G_c = g\right]$ indicates the monitor group g each municipality was classified into out of the total 3 G. The set of variables in $1\left[G_c = g\right] \times WindDir_{mt}^{90b}$ account for our excluded instruments, where each of the g clusters is interacted with $WindDir_{mt}^{90b}$ and equals 1 if the daily average wind direction in the municipality m is in the 90-degree interval [90b,90b+90) and 0 otherwise, with the interval [270,360) omitted (Deryugina et al., 2019a). Equation 4.4 is the first stage in which we expect $\hat{\beta}$ to be significant as an indication that wind direction at municipality m and time t affects the level of pollution at municipality m and time t (relevance assumption).

Following Deryugina et al. (2019a) and Bondy et al. (2020), we use k-means cluster to create 3 groups for all the pollution monitors in our sample. Clustering is largely used to assign pollution monitors into a pre-determined number of groups according to the groups' latitude and longitude. Hence, each group represents a geographic area [11]. To choose the optimal number of regional clusters, we use the k-means cluster algorithm (Makles, 2012). First, we follow the official IBGE regional division of the São Paulo and Rio de Janeiro States to define 11 the number of maximum possible clusters the municipalities in our

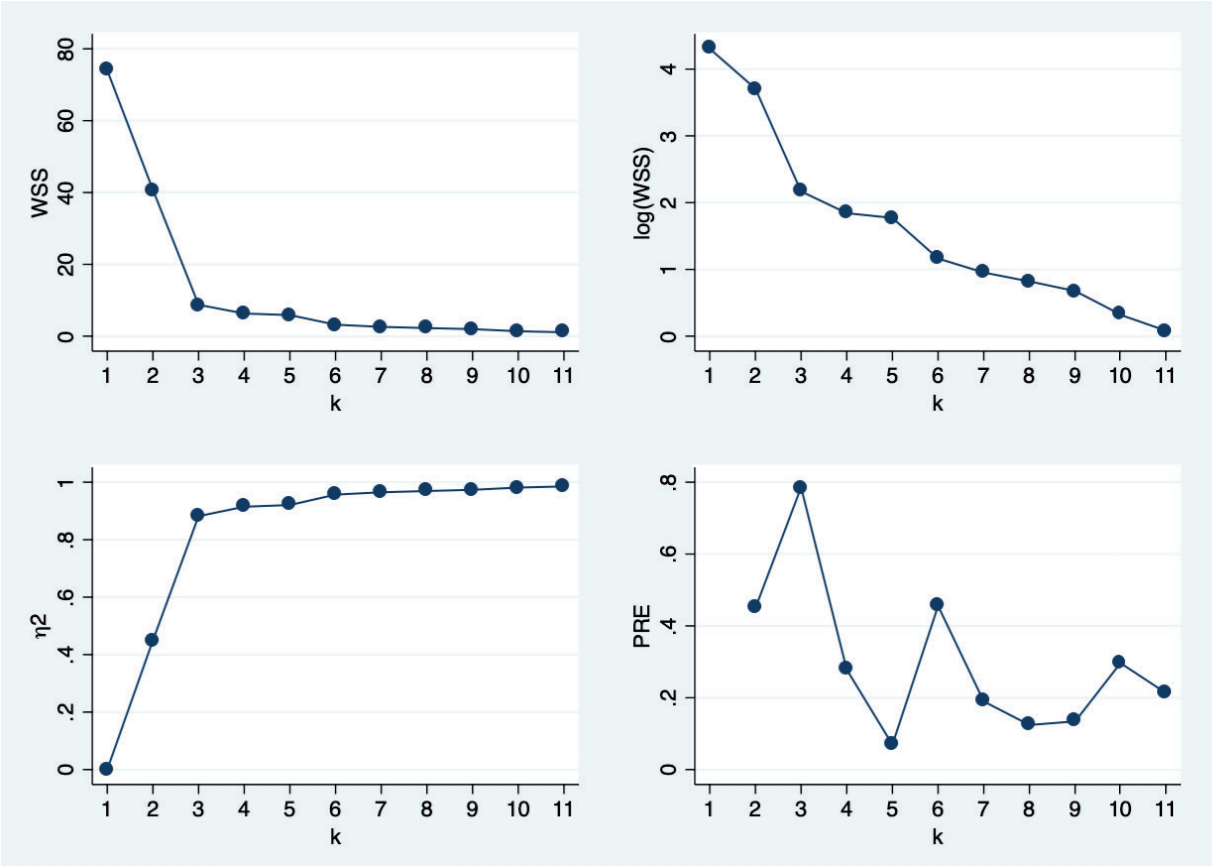
[11] K-means clustering aims to split n observations into k clusters in which each observation belongs to the cluster with the nearest mean. K-means algorithm minimizes within-cluster variances (squared Euclidean distances) and tends to find clusters of comparable spatial extent; i.e., by latitude and longitude in our case.

final sample could belong to. The IBGE adopts several forms to classify the territory into smaller regions, for example macro-regions, micro-regions and meso-regions. We adopted the latter classification in line with CETESB's and Inea's monitoring stations organisation to define the municipalities from our sample into those regional groups. We found 11 regional levels, which we inputed in the k-means algorithm to find the optimal number of clusters per latitude and longitude. Figure 3.1 shows the Within Sum of Squares (WSS), log(WSS), η^2, which is similar to the R^2, and the Proportional Reduction of Error (PRE) coefficient for all k cluster solutions. Accordingly, clustering with k=3 appears to be the optimal solution. More specifically, at k=3 we see a kink in the WSS and log(WSS). η^2 points to a reduction of the WSS by 88% and PRE to a reduction of about 79% compared with the k = 2 solution. Yet, the reduction in WSS is negligible for k > 3 [12].

Note that the number of groups are somewhat (by latitude and longitude) arbitrarily chosen (Deryugina et al., 2019a; Bondy et al., 2020). As a result, β_b^g varies across geographic groups (Deryugina et al., 2019a). It follows that, in using the interaction between the geographic groups with the wind direction dummies, we aim to mitigate the concern that a municipality's pollution monitors may not properly measure the average pollution each student within each municipality is affected by since the pollution station locations are sparse throughout the two states Rio de Janeiro and São Paulo. Furthermore, we also do not know the precise location of the exams venues within each municipality. As such, our identification strategy limits the effect of municipality wind direction on pollution to be the same for all stations within each of the 3 geographic groups, which comprise various municipalities. As explained by Deryugina et al. (2019a) and by Bondy et al. (2020), this confinement shortens the impact of pollution variation from different sources, reducing measurement error in our models. The concern here is that the 36 municipalities in our sample engulfs a large area and wind may carry particulate matter from different

[12] Our results do not leave doubt concerning the k=3 choice. Nonetheless, we k-means response depends on the the initial cluster centres chosen by the researchers. Thus, we repeated the process 50 times and found k=3 to be the optimal answer 83% of the times. Furthermore, for more details on how to calculate the WSS, η^2 and PRE, consult Makles (2012).

Figure 3.1: WSS, log(WSS), η^2, and PRE for all K cluster solutions (Optimal number of clusters



Note: The figure indicates clustering with k = 3 to be the optimal solution. At k = 3, there is a kink in the WSS and log(WSS), respectively. η^2 points to a reduction of the WSS by 88% and PRE to a reduction of about 79% compared with the k = 2 solution. However, the reduction in WSS is negligible for k > 3.

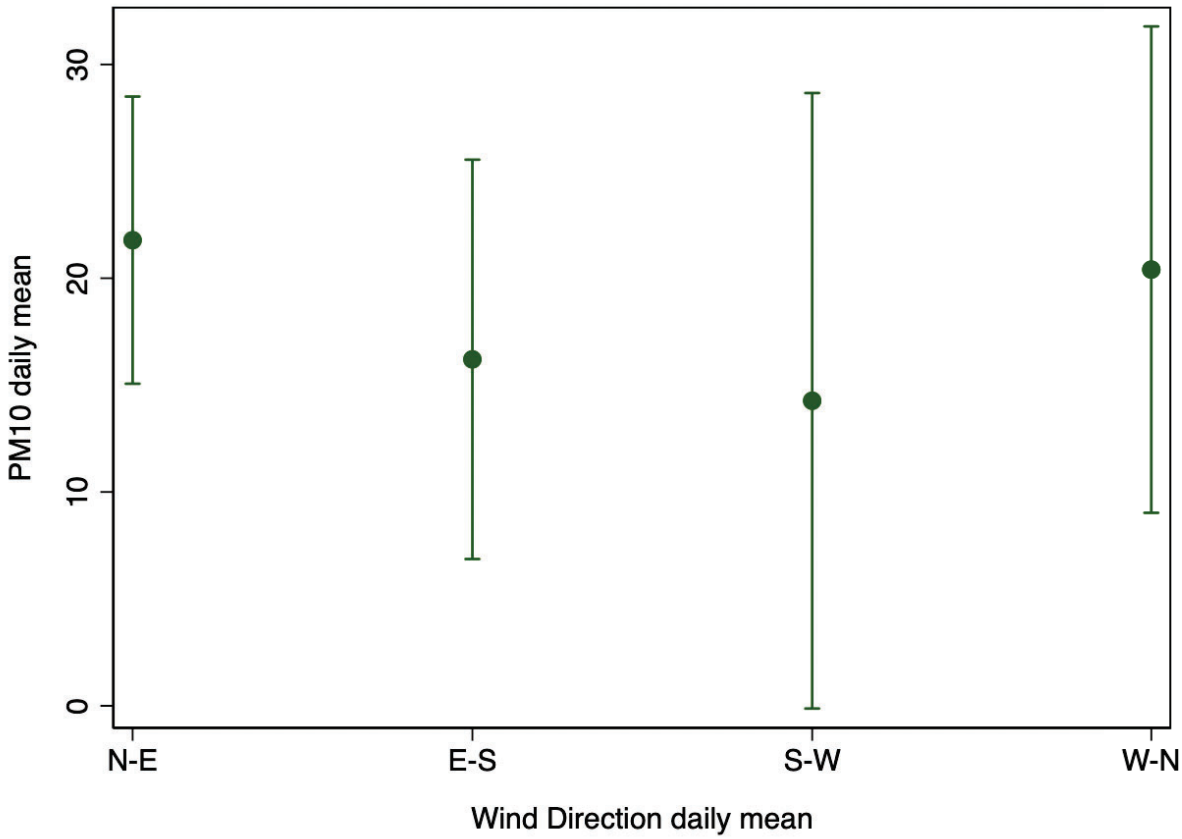
sources [13]. That is, if the station is close to the point source, its measures are likely to be higher than another station located far from the point source, which may cause measurement error in our estimations. On the other hand, nonpoint sources tend to present more similar influence on the monitors in the same group; and hence are more likely to be responsible for the variation on pollution that we expect to capture from the variables $1\left[G_c = g\right] \times WindDir_{mt}^{90b}$ (Deryugina et al., 2019a).

From Figure 3.2 shows the structure of our instruments by depicting the daily average PM_{10} across the 36 municipalities where the exams took place in our final sample by wind

[13]Note that we consider the direction the wind blows from, instead of the direction that the wind blows to as in Anderson (2015).

direction measured by its daily mean. It illustrates the relationship from our first stage suggesting that days with wind blowing from north, northeast and east have on average higher levels of PM_{10} [14].

Figure 3.2: The effect of wind direction on air pollution



Note: Daily average of PM_{10} concentrations for the 36 municipalities in the final sample per direction from where the wind blew during the exams days. 95% confidence intervals. Details on the data described in the text.

Our instrumental variable approach confirms our fixed effects results. Primarily, our valid set of instruments helps us to address potential unobserved factors that could remain in our baseline findings. Suffice it to say that our estimates are roughly the same in both empirical strategies. In addition, by using interactions between wind direction and geographical clusters, we are capable of demonstrating plausibly exogenous variation in air pollution measures across different regions comprising our sample.

[14]Table B1 shows the F test statistics for the strength of our instruments.

As for the validity of our instrument, wind direction fulfills the relevance requirement since it varies along with the concentration of PM_{10} as demonstrated by the F-test statistics in Tables B1 and 3.3. Moreover, with regards to the exogeneity requirement, the main assumption is that wind direction only affects students' scores via air pollution.

Although the previous literature has shown that O_3 and PM_{10} are not highly correlated, we use the daily average of pollution measures of PM_{10} as our main pollutant measure (Zweig et al., 2009; Ebenstein et al., 2016; Sager, 2019). This separate equations approach, i.e., the focus on one particular pollutant, has been used by, for example, Currie (2009) and Arceo et al. (2016). Other works prefer to use a composite measure of air pollution. [15] Nonetheless, we estimate equations with both pollutants jointly as a robustness check, and, as Table 3.6 shows, the estimate signs and magnitudes of O_3 and PM_{10} follow the same pattern as when PM_{10} is estimated in individual model.

For the temperature variables we build six bins of the size 3^0C, with the first bin comprising temperatures between 16^0C and 19^0C, and the last bin for temperature higher between 28^0C and 31^0C. This approach allows temperature to have a non-linearity impact. In all specifications, we use linear and quadratic terms for temperature (or bins) and humidity and their interactions (Ebenstein et al., 2016).

3.5 Empirical Results

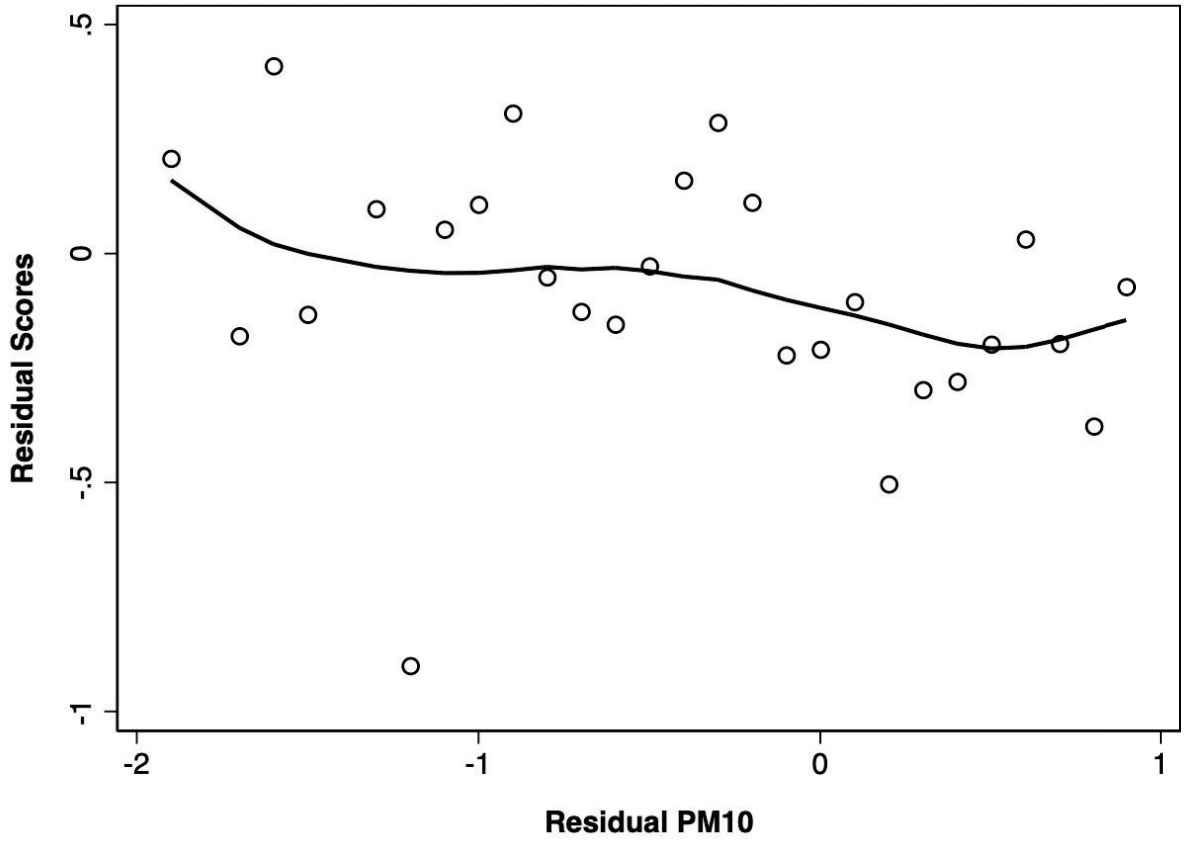
3.5.1 Main Results

Table 3.2 provides our estimation results. For completeness, we start by using municipality fixed effects. From Column 1 we can observe that an additional $10\mu g/m^3$ of

[15] For instance, Ebenstein et al. (2016), Chen et al. (2018) and Bondy et al. (2020) use the AQI.

PM_{10} corresponds to a reduction in students scores by 1%, although not significantly so. This result does not change after we include students' controls such as gender, race and mother's level of education. On the other hand, when it comes to utilising students fixed effects, the results in column 5 suggest that an increase of $10\mu g/m^3$ of PM_{10} on the day of examinations leads to a drop of 4% of a SD, though not significant when we do not control for the day of the exam. After controlling for day fixed effects, we notice, in column 7 - our preferred specification -, a highly significant reduction in students' scores of 8% of a SD for every $10\mu g/m^3$ increase in the same pollutant concentration.

Figure 3.3: Residual PM_{10} and Residual Scores Scatterplot (student FE)



Note: Plot of residual PM_{10} and residual Scores (SD) estimated by lowess bandsmoother. Residuals from a regression on students fixed effects and weather controls.

Figure 3.3 adds to our results in Table 3.2 by depicting that the negative relation between residual PM_{10} and students' scores is evident even when solely controlling for variation within students' idiosyncrasies. The graph is generated using the residuals from a regression of PM_{10} on students fixed effects and weather controls, showing that, on av-

erage, a student exposed to higher level of air pollution presents lower scores than another test taker exposed to better air quality during the exams.

Turning to our IV approach, the first stage's F-test statistic informs that wind direction is indeed a strong predictor of air pollution, hence, a strong instrument. Table B1 complements these results on the strength of our instruments. Our findings in Table 3.3 suggest that an additional $10\mu g/m^3$ of PM_{10} in the days of examinations caused the students' marks to drop by 8% (column 3). If we compare this with the baseline fixed effects results presented in Columns 1 and 2, we notice that the results remain pretty much the same even thought there is a slight reduction in the sample size.

To explore possible non-linearity between air pollution and students' cognitive performance, we present results from a complementary specification using a dummy variable for air pollution that equals 1 if the level of PM_{10} is above $20\mu g/m^3$ - WHO's acceptable threshold for this pollutant. Table 3.2 Columns 7 shows the estimates for our preferred model using linear pollution. One can note that the results are also significant with an increase of $10\mu g/m^3$ of PM_{10} causing a reduction of 11%SD in students' scores. When it comes to non-linear PM_{10} for the instrumental variable specification, Column 4 in Table 3.3 displays the same pattern of results as the one using linear pollution measure (column 3). That is, an additional $10\mu g/m^3$ on the day of the examination reduces students' scores by 13% SD. Also, regarding the first stage for the non-linear approach, the F test statistics still reveals that wind direction is a strong instrument for air pollution. These pieces of evidence suggest that cleaner air may lead to better educational outcomes and economic performance accordingly.

3.5.2 Heterogeneity across sub-populations

In this section, we analyse the heterogeneous effect of the treatment (air pollution) for our preferred specification reported in columns 7 and 8 from Table 3.2: (i) per gender;

(ii) across the distribution of scores per municipality; and (iii) in terms of different levels of income. Through this heterogeneity inspection, we aim to unveil how different sub-populations have their cognitive performance affected by the exposure to air pollution during high-stakes examinations.

With regards to gender differentials, our findings in Table 3.4 suggest that boys are more negatively affected by poor air quality during exams than girls. To put this into perspective, an increase of $10\mu g/m^3$ of PM_{10} causes males' scores to drop by 11%SD, whereas girls suffer a reduction of 6%SD. This heterogeneity is observed even when using a non-linear pollution variable; the results persist with boys' scores being reduced by 15%SD and girls' by 8%SD, meaning that boys who passed the examinations in days with air pollution levels above the WHO's acceptable threshold face a reduction in scores by 7pp deeper than girls (columns 6 and 7). This finding is consistent with previous studies including Jayachandran (2009); Sanders (2012), and Ebenstein et al. (2016). Although the role of gender in the effects of air pollution on respiratory health remains unclear, some epidemiology studies point out that males are more affected than females. For example, Abbey et al. (1998) link PM_{10} exposure to reduced lung function among males but not females, Galizia and Kinney (1999) find similar results regarding the exposure to O_3, and Chen et al. (2017b) identify that contemporaneous and cumulative exposure to air pollution has a stronger negative effect for men than for women. Similarly, Ebenstein et al. (2016) find the effect of pollution on exam performance to differ by gender and cite the increased incidence of asthma amongst boys as a possible cause.

As per the second heterogeneity analysis, our goal is to detect whether strong or middling students are differently affected by poor air quality. To this end we split our final sample into subsets per low and high scores within each municipality. One can note from Panel A Table 3.5 that students below the average have their scores 1pp more impacted than pupils that rank above the average; i.e., stronger students exposed to an additional

$10\mu g/m^3$ of PM_{10} during the exams have their performance reduced by 6%SD (Column 3), whilst weaker students are more negatively affected presenting a drop of 7% SD (Column 2). When we compare to the baseline estimation (column 1) using the whole sample, the average reduction is 8% SD. Another comparison is made between the top 5% scores and bottom 5%. Column 4 shows that students who rank very low in the exam suffer a reduction of 5% on their marks, whilst students on the top of the distribution seem not to be significantly affected by an increase of pollution levels during examinations (column 5). Finally, columns 6 to 10 bring results from the same exercises utilising the non-linear pollution variable. These findings suggest that students above the median and who took the exams in days with PM_{10} above the WHO'S threshold are more negatively impacted, with scores dropping by 12%SD, while students in the top 5% of the ranking are again not significantly affected by taking exams in days with pollution levels higher than $20\mu g/m^3$.

In Panel B Table 3.5, we slice our sample to look into how students from different socio-economic strata may be disproportionately affected by air pollution during the tests (Hsiang et al., 2019). For example, Neidell (2004) finds that less well-off households are more impacted by air pollution given their reduced capability of compensating behaviour. Similarly to the scores' distribution analysis, students from poorer backgrounds seem to be more affected by bad air quality. It may also be due to the fact that weaker students and low income are correlated Ebenstein et al. (2016). We therefore sub-sampled the individuals according to their household incomes and compared them using the mean of the municipality average income. From Panel B Table 3.5, one can notice that students below the median or below the bottom 5% face a reduction of 9%SD in their scores for every extra $10\mu g/m^3$ in the day of the tests (Columns 2 and 4), whereas wealthier students are not significantly impacted by a deterioration in air quality.

Table 3.2: The Effects of PM_{10} on Students' Scores (Main Specification with flexible weather controls)

	Municipality Fixed Effects				Students Fixed Effects			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PM_{10}	-0.01		-0.01		-0.04		-0.08***	
$(10\mu g/m^3)$	(0.03)		(0.05)		(0.03)		(0.03)	
1 if $PM_{10} > 20$		-0.04		-0.04		-0.09		-0.11***
		(0.03)		(0.03)		(0.07)		(0.04)
Students FE	N	N	N	N	Y	Y	Y	Y
Students Controls	N	N	Y	Y	N	N	N	N
Day FE	Y	Y	Y	Y	N	N	Y	Y
Weather controls	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,944,911	2,944,911	2,944,911	2,944,911	2,931,368	2,931,368	2,931,368	2,931,368
R-squared	0.004	0.004	0.17	0.17	0.05	0.05	0.07	0.07

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students controls includes gender, race and mother's level of education (columns 2 and 4), municipalities FE (columns 1 to 4) students FE, day FE (columns 5 and 6: there is no day FE), temperature in bins, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Table 3.3: Particulate Matters's Impact on Scores: Models with Instrumental Variable Wind Direction

Variables	(1) Baseline	(2) Baseline	(3) With Instruments	(4) With Instruments
PM_{10}	-0.08***		-0.08***	
$(10\mu g/m^3)$	(0.03)		(0.01)	
1 if $PM_{10} > 20$		-0.11***		-0.13***
		(0.04)		(0.01)
F First Stage			55.4	71
OBS	2,931,368	2,931,368	2,405,938	2,405,938

Note: See table 3.2 for full specification. Columns 1 and 2 identical to columns 3 and 4 from Table 3.2. Flexible weather controls; students and day FE. Standard errors clustered by municipality and bootstrapped with 500 repetitions.*** p<0.01, ** p<0.05, * p<0.1

Table 3.4: Heterogeneity in Particulate Matter's Impact on Students' Scores (in SD) per gender (with bins)

Variables	(1) All	(2) Girls	(3) Boys	(4) All	(5) Girls	(6) Boys
PM_{10}	-0.08***	-0.06***	-0.11***			
$(10\mu g/m^3)$	(0.03)	(0.03)	(0.03)			
1 if $PM_{10} > 20$				-0.11***	-0.08**	-0.15***
				(0.04)	(0.03)	(0.05)
Observations	2,931,368	1,671,910	1,259,458	2,931,368	1,671,910	1,259,458
R-squared	0.07	0.09	0.08	0.07	0.09	0.08

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE, temperature in bins, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Table 3.5: Heterogeneity in Particulate Matter's Impact on students' scores (in SD) across sub-populations (bins)

Variables	(1) Baseline	(2) Below 50%	(3) Above 50%	(4) Below 5%	(5) Above 95%	(6) Baseline	(7) Below 50%	(8) Above median	(9) Below 5%	(10) Above 95%
					Panel A: Low and high scores					
PM_{10}	-0.08***	-0.07***	-0.06**	-0.05***	-0.03					
$(10\mu g/m^3)$	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)					
1 if $PM_{10} > 20$						-0.11***	-0.07*	-0.09***	-0.05**	-0.04
						(0.04)	(0.04)	(0.03)	(0.02)	(0.03)
					Panel B: Low and high socioeconomic status					
	Baseline	Below 50%	Above 50%	Below 5%	Above 95%	Baseline	Below 50%	Above median	Below 5%	Above 95%
PM_{10}	-0.08***	-0.09***	-0.08**	-0.09***	-0.07					
$(10\mu g/m^3)$	(0.03)	(0.02)	(0.04)	(0.02)	(0.06)					
1 if $PM_{10} > 20$						-0.11***	-0.10*	-0.12***	-0.09**	-0.10**
						(0.04)	(0.05)	(0.04)	(0.04)	(0.05)

Notes: The dependent variable in all regressions is students' scores (in SD). Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE, temperature in bins, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

3.5.3 Sensitivity Analysis

To assess the robustness of our results, we experiment with a number of alternative specifications. So far we have reported results in which we include only one pollutant at a time in our estimations. Given the generally high correlation between many common local air pollutants there continues to be debate around whether or not different pollutants have independent health effects (see for example Lipfert et al., 2009; Katanoda et al., 2011; Cao et al., 2011; Gan et al., 2011; Chen et al., 2012). In our sample, O_3 and PM_{10} have a pairwise correlation of 0.10, significant at 5%. Nevertheless, Table 3.6 provides the results from estimations in which both pollutants are included together. In line with Bedi et al. (2020)'s work on cognitive performance in São Paulo, we see that PM_{10} continues to have the same impact on test takers (column 3), whilst O_3 does not present a significant effect on the latter. These are expected results since PM_{10} penetrates indoors easier than O_3, hence one would indeed expect lower or none impact of the latter pollutant in a co-pollutant estimation.

In a second exercise we include a one day lead in our main specification with the linear form of PM_{10} to capture common situations for example political events, weather shocks. We use data on air pollution from one day ahead of the actual date but which contains common features as the true measure. Table 3.7 shows that air pollution levels in the day after the exams do not significantly affect students' performance. Accordingly, an additional $10\mu g/m^3$ in the day of examinations is still responsible to a decrease of 5%SD in students' scores, while PM_{10} measures from the following day have null impact on pupils' marks. These results support our main findings not be driven by other factors that started before the test day and correlated with pollution, but solely by exposure to poor air quality on the day of the test.

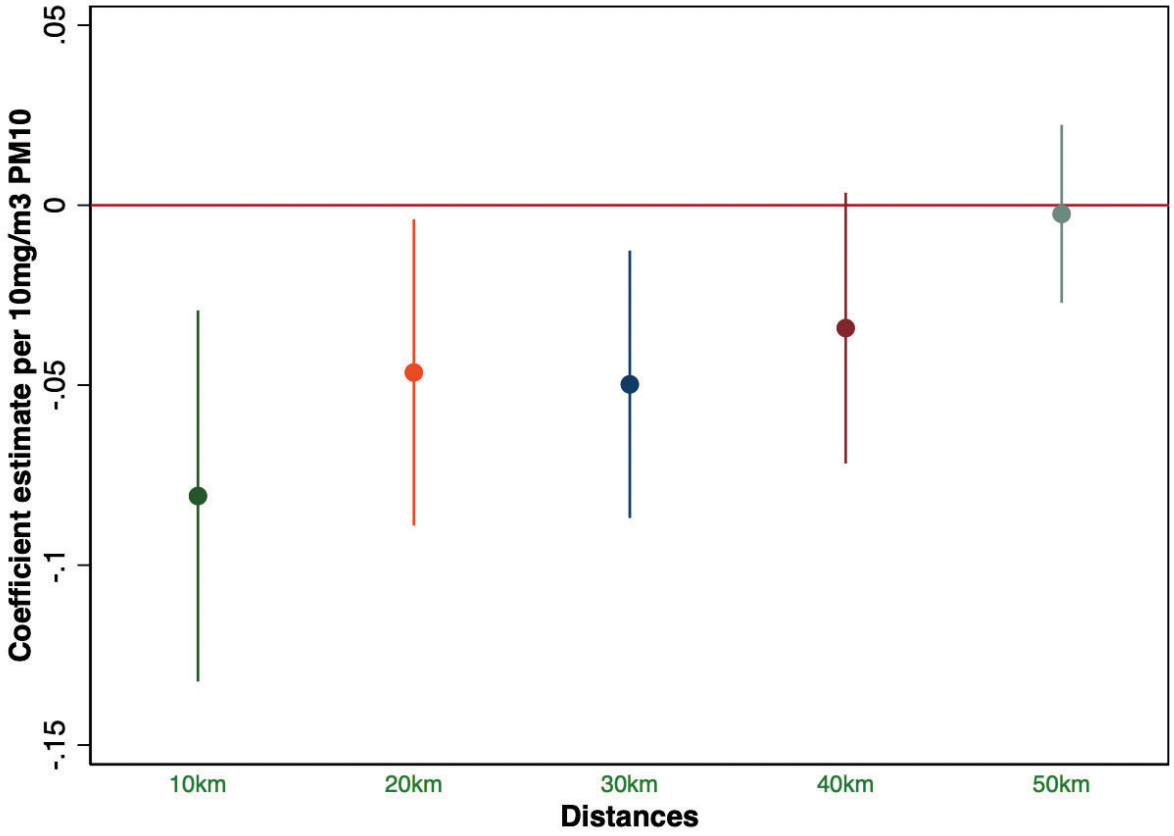
We next conduct a third exercise to check the robustness of our results. As we do not know the precise location where the students took their exams, we have so far used

the municipality centroid as a proxy for the exam venues and measured pollution using the average of up to 3 pollution monitoring stations within 10km of that centroid. However, if a school is located towards the periphery of a municipality then the monitoring stations within 10km of the centroid may not be the best measure of pollution levels at that school's location. Furthermore, we are omitting monitoring stations from our analysis which may be closer to the school's location but further away from 10km. To see whether this is affecting our results we here extend the 10km limit to 50km and plot the results in figure 3.4. We can see that PM_{10} continues to negatively affect boys' scores but the impacts fade and the standard errors inflate insofar as the distance rises, with the coefficients corresponding to the 50km estimation not statistically significant. This would seem to suggest that the 10km cut-off more accurately captures pollution concentrations at exam venues and that the pollutants' levels decay insofar as they travel [16].

Finally, to demonstrate that our main specifications with binned weather controls are not sensitive to different forms of meteorological controlling, Tables B3 and B2 present our main specification results without controlling for weather and with less flexible (with linear; and with squares and interactions controls). The results remain roughly the same, supporting our main estimates not be driven by unobserved weather conditions correlated with wind and students' performance, but solely by the effects of air pollution.

[16]Similar results can be found, for example, in Lai et al. (2018)

Figure 3.4: Cognitive impact of PM_{10} per distance from the station to the municipality centroid



Note: This figure reports the cognitive impact of air pollution within different distances. 10km is the baseline model. All the specifications follow the main model with students FE, day controls and flexible weather condition as in column 7 from Table 3.2. We note an increase of standard errors suggesting that the further the stations the less precise and significant the estimates.

Table 3.6: The Effects of PM_{10} on Students' Scores in SD (with co-pollutant (bins))

Variables	(1) Main model	(2) Main model	(3) With co-pollutant	(4) With co-pollutant
PM_{10}	-0.08***		-0.08***	
$(10\mu g/m^3)$	(0.03)		(0.03)	
1 if $PM_{10} > 20$		-0.10***		-0.05
		(0.04)		(0.04)
O_3			0.01	0.01
$(10\mu g/m^3)$			(0.02)	(0.02)
Observations	2,931,368	2,931,368	2,931,368	2,931,368
R-squared	0.07	0.07	0.07	0.07

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE, temperature in bins, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} and O_3 accounts for their linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Table 3.7: The Effects of PM_{10} on Students' Scores in SD (Baseline Model + 1 day lead)(bins)

Variables	(1) Baseline	(2) Baseline + lead	(3) Lead
PM_{10}	-0.08***	-0.05**	
	(0.03)	(0.02)	
PM_{10} (Lead)		-0.00	0.02
		(0.02)	(0.02)
Observations	2,931,368	2,172,680	2,172,680
R-squared	0.07	0.08	0.08

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Included students FE, day FE, flexible weather controls. PM_{10} is linear in $10\mu g/m^3$.

3.6 Conclusion

In this paper, we take advantage of the individual-level panel data structure of our data set to estimate the causal effects of contemporaneous exposure to air pollution on students' performance in university entrance exams from 2015 to 2017. To identify this relationship we use students fixed effects as our main specification strategy, and, as robustness checks, we utilise a wide range of control variables, such as individual characteristics, parents' level of education, household income, weather controls (temperature, humidity and wind speed), and day and municipality fixed effects. To address possible bias in our estimations stemming from classical measurement error and omitted variable bias, besides individuals fixed effects, we also perform an instrumental variable approach where we rely on the temporal and geographical variation in wind direction to build our instrument. Our main assumption is that this variation is exogenous since it does not directly affect the students' scores, except from its effect on air pollution.

We find that an increase of 10μ/m3 of PM_{10} in the day of the examination decreases the students' scores by around 6 points (8%SD). On top of that, when we study the response from different sub-populations within our sample to same levels of air quality, we find students from the bottom 5% of the economic ladder to be more responsive to a rise of 10μ/m3 of that pollutant, with a drop of 9%SD , whereas the 5% better-off exam takers present no significantly reaction to the exposure to PM_{10} in the day of the exams. Our results consistently suggest the harmful effect of air pollution on exam performance. Even when including a more flexible measure of our treatment; i.e., utilising a dummy variable to account for the days in which PM_{10} exceeded the WHO's acceptable threshold, our findings still point to the negative effects of air pollution on cognitive performance during examinations. Finally, we run several sensitivity checks and falsification tests, which suggest that our results are not driven by spurious correlations.

Two broad findings emerge from our results. Firstly, the effect of air pollution on

exam performance appears to differ for males and females, with the performance of the former more adversely affected. This finding is consistent with previous studies including Jayachandran (2009); Sanders (2012) and Ebenstein et al. (2016). Although the role of gender in the effects of air pollution on respiratory health remains unclear, some epidemiology studies point out that males are more affected than females. For example, Abbey et al. (1998) link PM_{10} exposure to reduced lung function among males but not females; and Galizia and Kinney (1999) find similar results regarding the exposure to O_3. Similarly, Ebenstein et al. (2016) find the effect of pollution on exam performance to differ by gender and cite the increased incidence of asthma amongst boys as a possible cause. Secondly, our results suggest that the effect of air pollution on exam performance differs by socio-economic background, with poorer students scoring even worse than wealthier treated exam takers. These findings are also in accordance with the literature, (see for example Neidell (2004) and Ebenstein et al. (2016), as there might exist a correlation between low parental income - and hence an impossibility of compensating behaviour - and pertaining to the bottom of the score distribution. All in all, our findings seem to add plausible evidence suggesting that cognitive performance may be hindered by bad air quality.

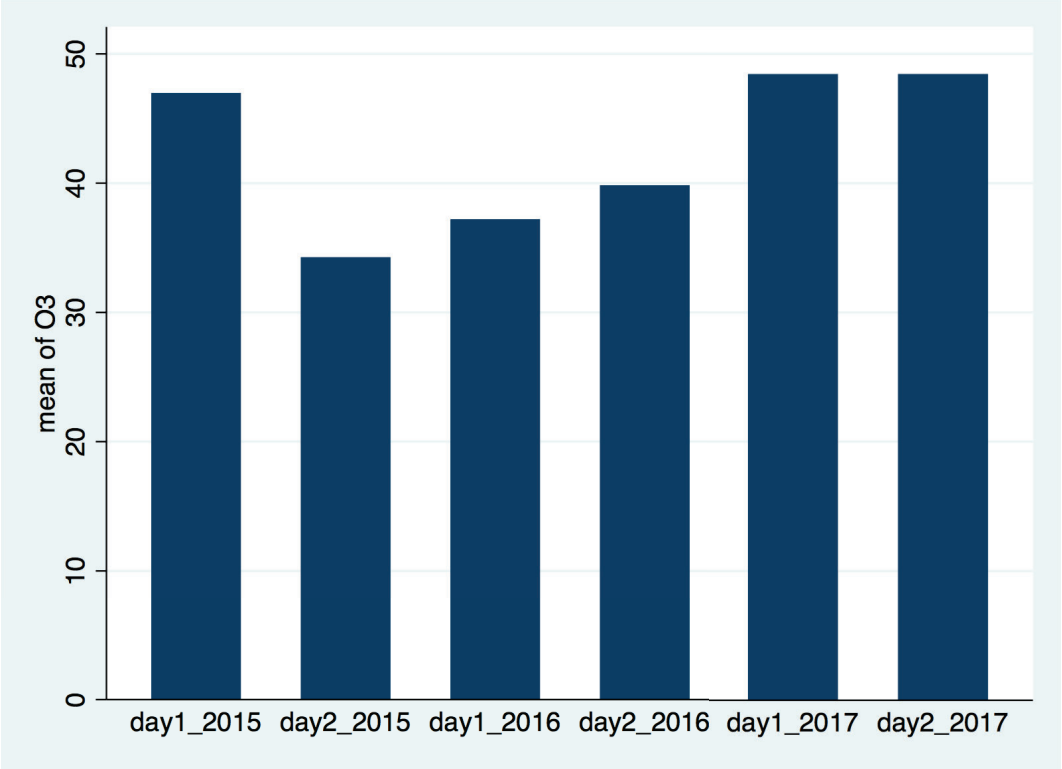
Appendix

Table B1: Instrument for Pollution: Wind Direction – First Stage Results

Cluster_direction	(1) PM_{10}	(2) $PM_{10} > 20$
clus1_90	0.00	0.41
	(0.63)	(0.24)
clus2_90	1.79***	1.43***
	(0.49)	(0.45)
clus3_90	-1.04**	0.12
	(0.44)	(0.3)
clus1_180	0.08	0.22
	(0.32)	(0.27)
clus2_180	1.3***	1.13***
	(0.38)	(0.36)
clus3_180	-0.26	0.62**
	(0.45)	(0.26)
clus1_270	-0.34	-0.02
	(0.23)	(0.2)
clus2_270	1.25**	0.97**
	(0.54)	(0.49)
clus3_270	-0.44	0.35
	(0.44)	(0.25)
Students FE	Y	Y
Day FE	Y	Y
F (First Stage)	55.4	71
Observations	2,405,938	2,405,938

Note: First stage estimates from main specification, using wind direction as an instrument to predict PM_{10} ($10\mu g/m^3$) in column 1; and for the dummy that takes 1 if $PM_{10} > 20$ in column 2. The instrument is the average wind direction in the day that the wind is blowing from each of 4 quadrants. There are 4 directions and 3 regions, hence we have (4-1)*3=9 instruments. Flexible weather controls. Standard errors clustered by municipality and bootstrapped with 500 repetitions.*** p<0.01, ** p<0.05, * p<0.1

Figure B1: O_3 mean per day - IWD - 10km



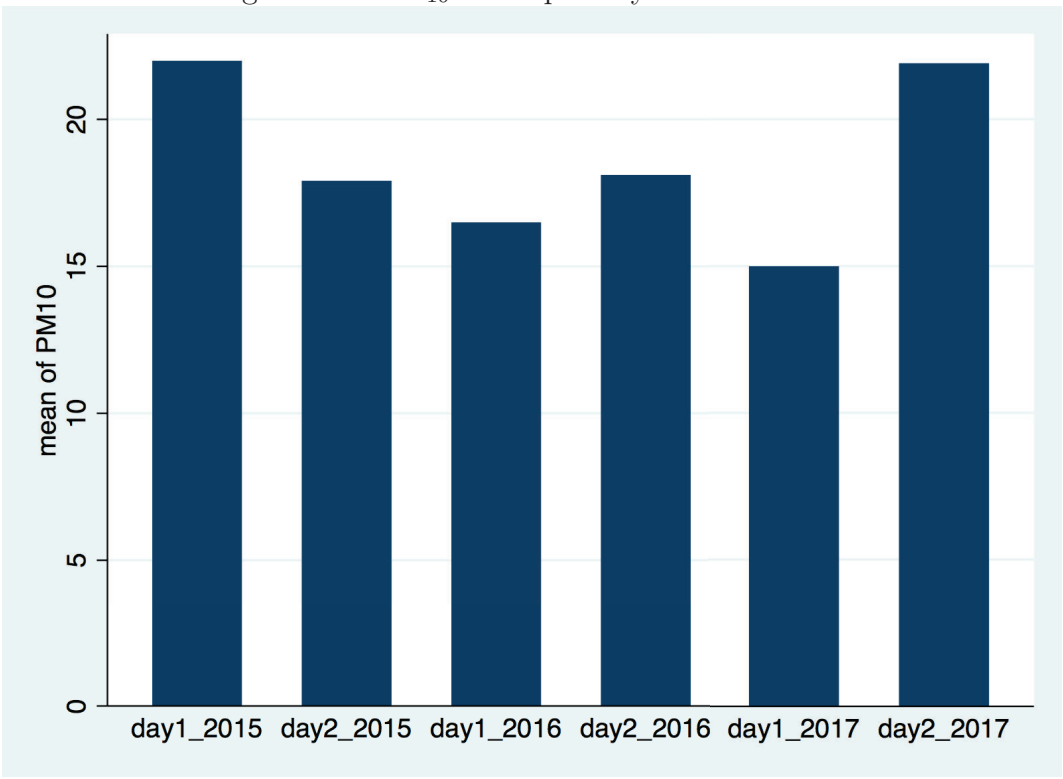
Note: This bar graph plots the average O_3 readings per day of all stations within 10 km from the municipalities centroids using the IWD.

Table B2: The Effects of PM_{10} on Students' Scores (Specification with flexible weather controls)

Variables	(1) Scores (in SD)	(2) Scores (in SD)	(3) Scores (in SD)	(4) Scores (in SD)
PM_{10}	-0.00		-0.03**	
$(10\mu g/m^3)$	(0.04)		(0.01)	
1 if $PM_{10} > 20$		-0.09*		-0.06*
		(0.05)		(0.03)
Students FE	Y	Y	Y	Y
Day FE	N	N	Y	Y
Weather controls	Y	Y	Y	Y
Observations	2,931,368	2,931,368	2,931,368	2,931,368
R-squared	0.04	0.04	0.07	0.07

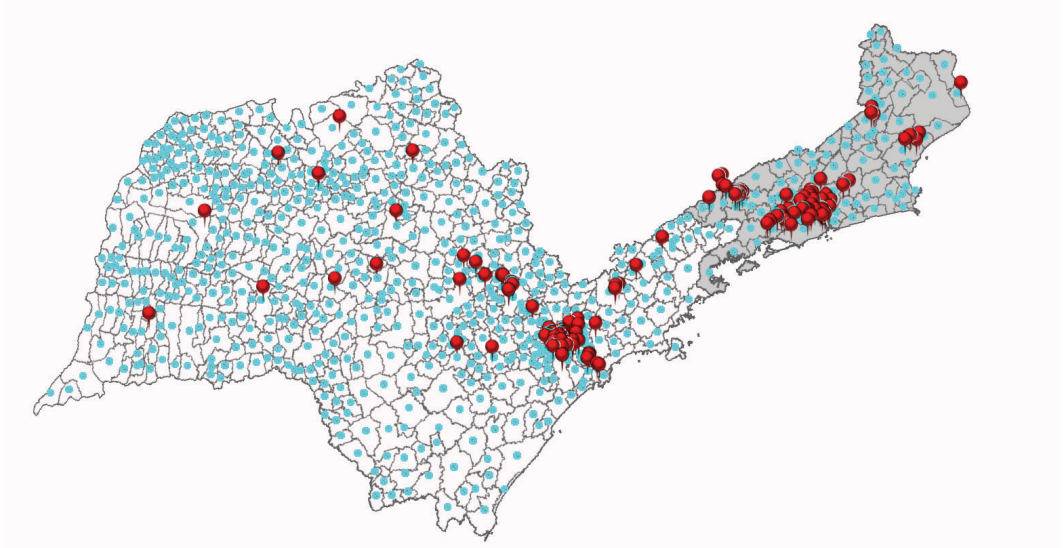
Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE (columns 1 and 2: there is no day FE), temperature and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Figure B2: PM_{10} mean per day - IWD - 10km



Note: This bar graph plots the average PM_{10} readings per day of all stations within 10 km from the municipalities centroids using the IWD.

Figure B3: Rio and São Paulo municipalities centroids and pollution stations (raw data)



Note: The figure plots Rio de Janeiro's (in grey) and São Paulo's (in white) maps at the municipality level. The circles represent the municipalities centroids; and the pins, the location of the pollution stations from our raw data set..

Table B3: The Effects of PM_{10} on Students' Scores (Linear and no weather controls)

Variables	(1) Scores (in SD)	(2) Scores (in SD)	(3) Scores (in SD)	(4) Scores (in SD)
PM_{10}	0.06**		-0.03	
$(10\mu g/m^3)$	(0.03)		(0.03)	
1 if $PM_{10} > 20$		0.08***		-0.04
		(0.03)		(0.04)
Students FE	Y	Y	Y	Y
Day FE	Y	Y	Y	Y
Weather controls	N	N	Y	Y
Observations	3,067,690	3,067,690	2,931,368	2,931,368
R-squared	0.03	0.03	0.06	0.06

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. Estimates for controls not shown: students FE, day FE, linear temperature, relative humidity and wind speed (columns 1 and 2: no weather controls). PM_{10} accounts for its linear form measured in $\mu g/m^3$; $PM_{10} > 20$ correspond to a dummy that equals 1 if $PM_{10} > 20$ and zero otherwise.

Table B4: Number of municipalities per distance from the monitoring stations

Distance (in km)	Municipalities	Observations
10	36	2,931,368
20	82	4,106,890
30	112	4,560,570
40	132	4,756,535
50	152	4,969,842

Notes: This table displays the number of municipalities and observations in samples that used 10, 20, 30, 40 and 50km distance between the closest monitoring station and the municipality centroid.

Chapter 4

Fetal exposure to fine particulate matter and students' cognitive performance: Evidence from agricultural fires in Brazil

4.1 Introduction

Controlled agricultural burning is a traditional technique still used by farmers, mainly in the developing world, to clean the field from previous harvests and prepare the soil for the next seeding phase. There is growing evidence, however, that the smoke from agricultural fires contributes to increasing levels of fine particulates in the air, causing, as a result, harmful health outcomes for nearby communities (Andreae and Merlet, 2001; Zhang et al., 2016; Chen et al., 2017a). As cognitive performance, via reductions in health, may also be harmed by air pollutants (Almond et al., 2018), investigating the impact of air pollution from these fires on cognitive performance and on future lifetime earnings is quite informative for the outline of agricultural and environmental policies. The purpose of this chapter is to estimate the long-run causal effects of fetal exposure to seasonal air

pollution from agricultural fires on nearby students' academic performance in the national examination in the 5^{th} grade in São Paulo, Brazil.

Despite a growing strand of research showing the short-term negative effects of agricultural fires on cognitive outcomes of the elderly (Lai et al., 2018), or its contemporaneous impacts in reducing students' scores at university entrance examinations (Graff-Zivin et al., 2020), to the best of our knowledge, this is the first study to quantify the long-run causal effects of prenatal exposure to air pollution from seasonal agricultural burning on pupils' scores and wages later in life. To fill this gap, we face a few challenges. Firstly, air pollution levels may be correlated with students' performance owing to their location choice (Bharadwaj et al., 2017) since parents may choose their residence according to the air quality of the region. If this happens, wealthier mothers will have their pregnancy in cleaner areas and, hence, their infants will be exposed to less air pollutants than children born to mothers living in more deprived and usually more polluted localities. Secondly, as seasonal agricultural fires increase both farmers' productivity and revenue, it may affect students' academic performance since the overall household income increases, allowing parents to invest more in their children. Thirdly, sugarcane is harvested in São Paulo in the winter, hence one would expect that some seasonality in weather conditions and winter-related diseases could raise spurious correlation between fires and fetal health. Although a number of studies in the medical literature show a negative effect of biomass burning on health (Arbex et al., 2004; Cançado et al., 2006; Ribeiro, 2008; Uriarte et al., 2009), or its positive impact on asthma hospitalisations (Arbex et al., 2007) in Brazil, these studies still fall short of showing a causal relation by failing to disentangle the confounding effects of other health determinants that are also correlated with controlled fires' smoke.

To tackle the endogeneity problem inherent in linking air pollution to health, and ultimately cognitive performance, we use the seasonal variation of agricultural burning

spots to assess the difference in air pollution. It turns out that, as explained above, fires can also be related to air quality measures. Thus, we follow Rangel and Vogl (2019) and interact fires with wind direction as the latter is able to predict spatial variation in the dispersion of smoke without being directly related to prenatal health or students' scores. As such, we are capable of causally evaluating the different cognitive outcomes of sugarcane burning on pupils living upwind and downwind from the fires. Our study adds to this previous work in three ways. First, Rangel and Vogl (2019) examine the short-term effects of prenatal exposure to agricultural fires on babies' health, while we look at the medium-run (pupils' scores in the 5^{th} grade) and at the long-term (wages). Second, we are interested in the effects of air pollution driven by those fires on human capital. That is, we show the mechanism through which fires affect human health. To this end, we adopt a different econometric approach by utilising, as an instrument, counts of fires upwind from the municipality's population centroid where the student attended school, which we consider being capable of generating plausible exogenous shocks to air pollution [1]. Finally, we use the fine particulate matter $(PM_{2.5})$[2], which are recognised as having a stronger negative link with health and human capital than PM_{10} used in previous work.

Our study is also related to Graff-Zivin et al. (2020), who investigate the effects of fires' smoke in the day of the Chinese university entrance examinations on candidates' performance, though we offer two more contributions as we pay heed to the medium and long-terms effects of air pollution, while they look only at the contemporaneous impacts. In addition, we complement their study as we are able to establish the causal relation between air pollution and students' cognitive outcomes, an aspect that Graff-Zivin et al. (2020) are not able to examine due to a lack of pollution data for the exact moment of the exams. Also, regarding gender heterogeneity, our work provides evidence that boys' cognitive performance is more impaired by poor air quality than girls', again an aspect that

[1] Unfortunately, we do not have data either on the exact place of birth or on migration. Therefore, we use each municipality's population centroid as a proxy for municipality of birth.

[2] $PM_{2.5}$ refers to fine aerosol particles with a diameter of less than or equal to 2.5 micrometers (μm) near the surface.

Graff-Zivin et al. (2020) lacked data for. More generally, this chapter adds to an extensive body of literature that relates the impact of air pollution to negative health outcomes (Seaton et al., 1995; Chay and Greenstone, 2003; Currie and Neidell, 2005; Ebenstein et al., 2017); to a decrease in productivity of low and high-skilled workers' (Chang et al., 2016; Heyes et al., 2016; Archsmith et al., 2018; Chang et al., 2019; Kahn and Li, 2019); to an increase in school absenteeism (Currie et al., 2009; Ransom and Pope, 2013; Chen et al., 2018; Liu and Salvo, 2018); and to a decrease in cognitive performance (Currie et al., 2009; Zweig et al., 2009; Reyes, 2011; Ebenstein et al., 2016; Bharadwaj et al., 2017) due to its harmful effects on the respiratory and cardiovascular systems (Dockery et al., 1993; Seaton et al., 1995), insofar as we aim to demonstrate the medium and long-run damages caused by air pollution derived from agricultural fires during pregnancy in locations where $PM_{2.5}$ measures are not as high as in the vast majority of other studies. In addition, our pollution source is not only delimited but also within an agricultural setting, differing from other already explored sources, such as vehicles, plants, chimneys, or airports; thus, expelling different chemical substances to the atmosphere. To put it into perspective, we calculate that around 61% of fires take place in agricultural lands within our data set[3]. As such, we shed light on the important fact that, even at relatively moderate levels, this pollutant has the potential to cause cognitive ability impairment.

Some studies have shown that exposure to Ozone, and to the pollutants it stems from (NOx), is related to premature death, asthma, bronchitis, heart attack, and other cardiopulmonary illnesses (Weinhold, 2008; Currie et al., 2009; Neidell, 2009; Developers, 2016; Wikipedia, 2019), and, even at low levels, more susceptible individuals are more likely to be affected by this pollutant (Bell et al., 2006; Weinhold, 2008). Regarding SO_2, this pollutant has a number of harmful health impacts, including breathing difficulty, pulmonary edema, eye irritation, asthma attacks, cardiopulmonary diseases and increase of

[3]The rest is spread across urban areas (3.2%), grazing(15%), forestry(15.3%), with the remaining fires being spotted in forests and other fields. To make these calculations, we overlay our fires data from NASA/FIRMS with land coverage data from the agricultural census from 2010 available at the IBGE website.

mortality rates (Lin et al., 2004; Khan and Siddiqui, 2014; Goudarzi et al., 2016; Khaniabadi et al., 2017). As the brain requires large amounts of oxygen to function well, medical studies have suggested that the cognitive performance is likely damaged as a result of the exposure both in the short and long-run to air pollutants (Calderón-Garcidueñas et al., 2008). The medical literature shows that different parts of the brain perform different functions, with white matter being linked to language and verbal reasoning skills and gray matter associated with mathematical skills (Popescu et al., 2019; Houston et al., 2019). Other works associate exposure to air pollution with a decrease in total gray and white matter volumes (Erickson et al., 2020). Based on this evidence, the economics literature has suggested a causal link between pollutants and a deficit in cognitive function (Currie et al., 2009; Zweig et al., 2009; Reyes, 2011; Ebenstein et al., 2016; Bharadwaj et al., 2017).

We focus our attention on sugarcane since the bulk of the fires take place in areas dedicated to this crop. Nonetheless, there are others crops that also use the burning technique, such as maize, rice and wheat. One can observe from Figure C8 the share of sugarcane production per total cultivated land in all municipalities in São Paulo and, comparing to Figure C9, sugarcane plantation has increased since 2000, shifting the other three products away. Furthermore, when it comes to relating the occurrence of fires with sugarcane, we can notice from Figures C1 and C3 that the areas with more fires detected coincide with the hot spot areas of this crop. Sugarcane fires take place in large areas emitting aerosols, such as poly-cyclic aromatic hydrocarbons (PAHs), organic and black carbon, sulfur dioxide (SO_2), ozone (O_3), nitrogen oxides (NOx) and particulate matters ($PM_{2.5}$ and PM_{10}) and is carried by wind towards surrounding localities, which usually are very populated (Guoliang et al., 2008; Akagi et al., 2011; Wu et al., 2018). This feature makes straw burning a rather suitable source of air pollution given that not only do we know the exact location of the fire points, but also the fact that agricultural fires are close to populated regions, raising concerns about human capital damages much more than if we were exploring wild fires, which generally take place in remote lands. Last

but not least, this chapter complements Bharadwaj et al. (2017) and Isen et al. (2014) in calculating the returns of cleaner air quality on scores and, consequently, on wages (long-run effects) in the largest Latin American country.

With regards to our empirical design, we build an individual-level pooled cross-sectional data set, collected from the Ministry of Education of Brazil, with scores from students born between 2001 and 2008, which we merge by month, year of birth and municipality of residence with the exact location of the active fires. The latter data come from the National Aeronautics and Space Administration (NASA)'s active fires product. Besides that, we use daily atmospheric conditions from satellite remote sensing re-analysis gridded data by the Japanese 55-year Re-analysis (JRA-55), from which we utilise the average wind direction per day with or without fire to create our instruments. These consist of the counts of fires per wind direction: upwind, downwind and vertical wind. That is, we consider upwind fires those whose smoke would blow towards a certain municipality's population centroid; downwind fires are those in the opposite direction; and vertical are fires in all the remaining directions. Our first stage results show that, under our instrumental variables approach, our main pollutant measure, $PM_{2.5}$, can be considered randomly assigned since we find stronger and more significant correlation between upwind fires and pollution than other fires. In addition, our instruments of fire counts are built in several ways, and proven to be strong for $PM_{2.5}$. To link students' scores to air pollution level *in utero*, we use satellite gridded data from NASA's second Modern-Era Retrospective analysis for Research and Applications (MERRA-2) product, merging it by municipality-month-year of birth.

Our results indicate the external costs borne by individuals stemming from agricultural burning inasmuch as it impedes the capital formation and, ultimately, the economic growth of the country (Solow, 1999). This is of particular importance when it comes to developing countries, where the returns to education are generally higher than developed

nations, making São Paulo an interesting case study since the state yields around one fifth of the global production of sugarcane, with the largest production within the country and in the world (FAOSTAT, 2020; Unicadata, 2020). In response to a national and international rise in demand for bio-fuel and sugar, the state more than doubled from 10% to 23% (Figures C1 and C3 offer a visualisation of this) its share of area designated to sugarcane production throughout the past two decades, containing two thirds of all Brazilian production (McConnell et al., 2010; Rangel and Vogl, 2019). Moreover, despite the relative decline in the number of agricultural fires in the past years in Brazil, due to become illegal from 2021 with the Green Ethanol Protocol in 2017 signed by the government (SIMA, 2017), controlled burning is still a technique substantially applied in the country, with burning crops starting to see its peak in our period of study as Figures C5 and C7 illustrate. The harvesting season for sugarcane starts in April until November, when the farmers who use the manual method - instead of the mechanical - set controlled fire to the fields to crop the cane, increasing the labour productivity had the method not been applied (Fernandes, 1988; Kirchhoff et al., 1991; Rangel and Vogl, 2019). Usually the fires take from a few hours to 24 hours during the peak (Lamsal et al., 2017).

In this setting and consistent with existing research, our findings suggest that an increase of 10μ/m3 of $PM_{2.5}$ during pregnancy decreases students' Portuguese scores by 4.6% and Maths scores by 2.8%, with boys more affected than girls. Back-of-the-envelope calculations indicate that a reduction of $PM_{2.5}$ in the third trimester by 2.4 SD, i.e., 10μ/m3, would increase child's scores by 4.6% and ultimately her wages by 2.3%. Our results are robust to a set of sensitivity checks. Therefore, our study provides policymakers with important evidence of a link between prenatal exposure to air pollution and lower cognitive performance.

The remainder of this chapter is organized as follows. In Section 2, we describe the data. In Section 3, we detail the empirical strategy. In Section 4, we present results. In

Section 5, we show sensitivity analysis, and Section 6 concludes.

4.2 Data

We build a repeated individual-level cross-sectional data set on monthly air quality, weather conditions, agricultural fires and students' scores in the fifth grade in São Paulo. We choose to examine cohorts of students born between 2001 and 2008 because some of our remote sensing data set come from the satellites Terra and Aqua, which started to operate between 1999 and 2002. Furthermore, we opt for remote sensing data instead of ground-based monitoring stations since the agricultural fires are eminently located within rural areas, in which very scarce pollution and meteorological stations could be found for our time period of analysis. Our data are collected from four main sources as detailed below.

4.2.1 Agricultural Fires

We collect daily remote-sensing data on agricultural fires from the National Aeronautics and Space Administration (NASA)'s Fire Information for Resource Management (FIRMS). The data are captured by two satellites TERRA and AQUA, which rely on Moderate Resolution Imaging Spectroradiometer (MODIS) sensors to detect ground-level fire activity. Each satellite overpasses Brazil twice per day: in the morning, afternoon, evening, and late nights (Kaufman et al., 1998; Justice et al., 2002). The fire detection algorithm developed by researchers at the University of Maryland is built upon thermal anomalies, surface reflectiveness, water, cloud masking and land use, and the publicly available data is provided with a resolution of 1km X 1km (Giglio et al., 2009, 2016). Notwithstanding, it is not possible to know the precise size of each fire from the data.

The product also gives the confidence of detection for each potential fire identified by the satellites. That is, each fire receives a probability of real occurrence (from 0% to 100%) according to the meteorological conditions and vegetation at the exact time the fire was detected (Giglio et al., 2016). As a result, in our specifications we use the fire weighted by its confidence to minimise classical measurement error (Rangel and Vogl, 2019; Graff-Zivin et al., 2020).

A fire is attributed to a student if it occurred within a catchment area of 50km radius of her municipality's population centroid. Unfortunately, we do not know the municipality of birth of each individual. Hence, we use the municipality's population centroid where the pupil lives when she takes the exams as a proxy for the exact location of birth. Nevertheless, alternative distances are used and presented in the robustness checks section.

To build our fire variables, we follow Rangel and Vogl (2019) and count the daily number of grid cells containing a fire within a 50km of a municipality's population centroid. Then, we consider that fires detected within the same wind octant in a day are upwind, fires located in the opposite octant are downwind, and all the remainder fires are defined as vertical. As in previous works (Rangel and Vogl, 2019; Graff-Zivin et al., 2020), for our main specification, we set a 45^0 angle. As such, we define upwind fires as those located within a 45 central angle from the average daily wind direction within each municipality. In other words, upwind fires have the smoke blowing towards the municipality's population centroid. Downwind fires are located in the opposite direction; and we call vertical fires the ones situated in the other directions. The rationale of this classification is that municipalities situated upwind from fires are exposed to more air pollution coming from the fires than downwind locations, whereas the air pollution in the municipalities located in the vertical direction is less influenced by smoke than the upwind villages, but more affected by fires than their downwind counterparts. In our sensitivity analysis section we aggregate vertical and downwind directions to form an alternative set of instruments

since these directions present similar behaviour. Moreover, we run models using upwind and downwind fires, excluding the vertical ones, and the results remain very close to the main specification.

In Table 4.1, we report the summary statistics for agricultural fires per wind direction used for the whole period of the main estimation sample (2001-2008). Also, Figures C5 and C6 present the time series behaviour of monthly average fires for that period. From these figures we can observe a strong seasonal pattern of fires with peaks in the winter months, and a not very clear time trend. The seasonality of fires coincides with the agricultural phases, corroborating the link between sugarcane harvesting period and the peak in identified burning spots. In Figure C7, we present the overall time series behaviour from 2001 until 2017; i.e., we add the period ranging from 2009 to 2017 since the students took exams from 2011 to 2017.

4.2.2 Atmospheric Conditions

Re-analysis Data

We use the Japanese 55-year Re-analysis (JRA-55) data on atmospheric conditions. The JRA-55 is the second Japanese global atmospheric re-analysis project conducted by the Japan Meteorological Agency (JMA), covering data back to 1958.

Re-analysis uses past climatic observations from satellite and terrestrial monitoring stations, as well as meteorological models to systematically produce data sets for climate monitoring and research. On one hand, the main advantages of working with re-analysis data are: i) their consistent spatial and temporal resolution and baseline models which have reduced biases over each product's new edition; and ii) re-analysis inputs millions of observations that would be impossible for a researcher to collect and deal with it separately. On the other hand, the key limitation is the biases in observations and models that

lead to unavoidable classical measurement error. Hence, the re-analysis data and results of models using them should be interpreted with caution.[4] Nevertheless, to attest the scientific validity of this data set, several works in the Atmospheric Sciences field evaluated the precision of the JRA-55 products in comparison to others using different models by different researchers, finding that these re-analysis products lack of perfect accuracy but are still useful for climate studies (see for example, Sun et al., 2018; Tsujino et al., 2018).

Weather Conditions

We gather gridded data on temperature in Kelvins, which we transform to degrees Celsius (^0C), precipitation (in millimeters, mm), relative humidity (in percentage, %), wind components and cloud coverage (in percentage, %). Apart from the latter, these weather conditions may influence the way air pollution geographically spreads. As such, we include them as control variables in our models. Regarding cloud coverage, it is important since the satellite imagery for fire detection is negatively related to the amount of cloud in the sky.

The spatial resolution of each grid is 0.56° X 0.56°, and the temporal is 3-hourly data that we average by day per grid. We then calculate the monthly average at the grid level per occurrence of fire. That is, we find the monthly mean by aggregating the days with fires and the days without registered fires since weather conditions may different in days with and without fires. By doing so, we aim to mitigate classical measurement error had we imputed weather conditions from days without fires for days with fires and vice-versa.

To construct the variables for wind conditions we use the East-West wind direction (u-component) vector and the North-South (v-component) vector. The wind speed (in

[4]For more details on how re-analysis data are produced, please, visit the website https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables.

meters per second, m/s) is calculated by:

$$Wind\ speed = \sqrt{u^2 + v^2} \tag{4.1}$$

To calculate the average wind direction for each observation, firstly, we transform the degrees into radians; secondly, we compute the North/South direction = y = cosine(radians) X wind speed; the East/West direction = x = sine(radians) X wind speed; sum all ys (Y) and sum all xs (X); calculate the arctangent (the inverse tangent) using Y and X. The results are finally transformed back from radius to angle (Webmet, 2019; Itron, 2019). One should note that the average wind direction refers to the direction the wind is blowing from since we are interested in the direction of the wind in relation to the fire point blowing towards the municipality's population centroid.

4.2.3 Air Pollution

The pollutants and other chemical elements, such as $PM_{2.5}$, SO_2, Sulfate (SO_4), O_3, Black Carbon (BC) and Organic Carbon (OC), dust (DS) and sea salt (SS), utilised in this chapter come from the second Modern-Era Retrospective analysis for Research and Applications (MERRA-2), provided by NASA and built upon a plethora of data sources, such as remote sensing imagery and terrestrial stations. The gridded data available has a spatial resolution of 0.5° X 0.625° and comes in hourly intervals, which we aggregate to daily and monthly frequency. To this end we implement the same procedure as for the weather variables; i.e., we first average per day distinguishing days with or without fires, then we calculate the monthly average.

$PM_{2.5}$ is not directly available from the Merra-2's data set. Thus we follow the Atmospheric Science literature to calculate the particulate matter using the formula below

(Buchard et al., 2016; Provençal et al., 2017).

$$PM_{2.5} = 1.375 * SO_4 + 1.8 * OC + BC + DS + SS \tag{4.2}$$

where SO_4, OC, BC, DS and SS are measured in kilogram per cubic meter (kg/m^3), which we finally transform to micro-gram per cubic meter (μ/m^3).

As per any re-analysis data, one needs to be careful when interpreting our results using these pollutants data since their use feasibly could induce some attenuation bias in the econometric estimations. More precisely, measuring pollution via remote sensing still present a few weaknesses which have been assessed by the Atmospheric scientists (Liu et al., 2009; Paciorek and Liu, 2009; Estes Jr et al., 2010). For example, Buchard et al. (2016) and Provençal et al. (2017) assess the precision of satellite-based aerosol optical depth (AODS) in comparison to air pollution monitored on the ground in the United States and in Europe. Both works demonstrate that the AOD figures are quite reliable in relation to the terrestrial stations' readings.

4.2.4 Education

Our data on students' scores come from the National Institute for Educational Research *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (INEP), an agency from the Brazilian Ministry of Education. Every two years the students in the fifth grade pass a nationwide examination to assess the quality of education in Brazil (*Prova Brasil)* via tests and socioeconomic questionnaires (INEP, 2017a). This forms a rich administrative data set with detailed information on the students' scores in Portuguese and Mathematics, demographic information, such as parental level of education, age, race, gender and special needs. From this data we build a repeated pooled cross-section at the student level with cohorts taking examinations from 2011 to 2017 (Table

4.1 summarises this information). The Portuguese scores range from 0 to 350 points, and Maths from 0 to 400 points, and, in our regressions, we follow the literature and use these scores in standard deviation form as dependent variables (Garg et al., 2020). In addition, we include cohorts of students taking the fifth grade exams in 2011, 2013, 2015 and 2017. As for the year of birth, the children in our data set were born from 2001 to 2008.

The INEP data provide the highest level of education achieved by each student's parents. As such, we first drop the missing observations in which the students declared they were unaware of their parents' level of education. Then, we create six dummies for each parent's education: (i) never studied, (ii) incomplete 1^{st} elementary if the parent attended school for a while, but did not complete the first elementary, (iii) incomplete 2^{nd} elementary if the parent attended school for a while, but did not complete the second elementary, (iv) elementary if the parent only completed elementary School, (v) high school if the parent completed the high school, and (vi) undergraduate college education (Zweig et al., 2009). To avoid the dummy trap, we exclude one dummy for each parent's education. We also control for students' race and gender.

In Table 4.1, we present the descriptive statistics for our main estimation sample with 1,190,717 students. The average daily concentration of $PM_{2.5}$ in our estimation sample is 14.3 $\mu g/m^3$, with a standard deviation of 4.24. For SO_2 the respective figures are 4.06 $\mu g/m^3$ and 2.82.

Table 4.1: Summary Statistics for main regressions sample

Variables	Mean	SD	Min	Max	Obs
Scores					
Portuguese	219	47.8	84.9	339	1,190,717
Maths	234	47.5	78.92	367	1,190,717
Pollutants					
$PM_{2.5}$ ($\mu g/m^3$)	14.3	4.24	4.64	59.5	1,190,717
SO_2 ($\mu g/m^3$)	4.06	2.82	0.11	28.8	1,190,717
Agricultural Fires					
Upwind	8.61	12.8	0	129	1,190,717
Downwind	7.49	10.5	0	125	1,190,717
Vertical	14.5	20.8	0	191	1,190,717
Meteorological Conditions					
Rain (mm)	2.73	2.66	0.00	92.4	1,190,717
Temperature (^0C)	22.2	2.69	13.4	35.6	1,190,717
Wind Speed (m/s)	2.05	0.77	0.13	8.52	1,190,717
Cloud coverage (%)	40.4	13.1	6.24	99.9	1,190,717
Humidity (%)	70.3	8.54	25	97.6	1,190,717

Notes: Summary statistics of some of the key variables: scores ranging from 0 to 350 points for Portuguese and 0 to 400 points for Maths, counts of agricultural fires and meteorological conditions during pregnancy (from 2000 to 2008) in the municipalities in the State of São Paulo. The fires locations are defined within 45 degrees from the daily average wind direction in a municipality.

4.3 Empirical Strategy

4.3.1 The basic OLS models

We examine the partial correlation between $PM_{2.5}$ fetal exposure and test scores in the 5^{th} grade by pooling our individual-level cross-sectional data for the exams happening in years 2011 to 2017. The basic specification takes the following form:

$$Y_{jmit} = \phi pollutant_{mi} + \psi W_{mi} + \alpha X_j + \mu_m + \iota_i + \theta_t + \epsilon_{jmit} \tag{4.3}$$

where Y_{jmit} is the 5^{th} grade test score in either Portuguese or Maths of child j born in municipality m, at time i, taking exam at year t. The parameter of interest is ϕ, i.e., the coefficient of $PM_{2.5}$ measuring the average level of pollution at the trimester and

municipality levels. Our study focuses on the effect of $PM_{2.5}$ on cognitive performance, although in robustness exercises we include SO_2 as a co-pollutant. W represents a vector with weather conditions controls during pregnancy, such as temperature, relative humidity, wind speed, cloud coverage, and rainfall per trimester. Since temperature extremes can affect fetal health and air pollution concentration, suggesting non-linearity in its effects, we create five temperature bins between less than 15 and more than 35 degrees (5 degree bins) in our main specification, but also, for completeness, experiment with using temperature in the linear form 4.8, squared, and interacted with other weather variables (Deschênes et al., 2009; Bharadwaj et al., 2017). X_j is a vector with controls for observable individual characteristics that can be related to their exam scores, such as the mother's highest level of education[5], students' race and gender. μ_m are municipality fixed effects, we include ι_i, a vector to control for month and year of birth fixed effects. The two latter account for unobserved time-invariant municipality and cohort characteristics, and θ_t contains dummies to control for year of exam. The standard errors are clustered by municipality to control for spatial and serial correlation within each locality. Unfortunately, we do not have the exact place of birth, thus we use the municipality where the student attended school as a proxy for municipality of birth.

Some concerns arise in using OLS. The first relates to the fact that students' exam scores may be correlated with air pollution since the wealthier parents could choose to live in less polluted areas (Banzhaf and Walsh, 2008; Currie et al., 2014). This location choice is a form of avoidance behaviour, which may bias our estimations and needs to be tackled. (Neidell, 2004, 2009; Bharadwaj et al., 2017). Second, as seasonal agricultural fires increase both farmers productivity and revenue, it may affect students' academic performance since the overall household income increases, allowing parents to invest more in their children. Third, sugarcane is harvested in São Paulo in winter, hence one would

[5]We do not have the years of education of the student's parents. Rather, the original micro-data provides the highest level of education achieved by them. As such, we build dummies indicating each of these levels. For more details on the construction of the dummies, please refer to section 3.

expect that some seasonality in weather conditions and winter-related diseases could cause a spurious correlation between fires and fetal health. To try to mitigate these challenges we opt by using an instrumental variable approach to estimate our models. The novelty of our work is that we combine agricultural fires with wind direction to predict air pollution.

In addition, our type of pollution measure and meteorological data set, which are produced by re-analysis modelling, may bias our estimates inasmuch as children's actual exposure to outdoor air pollution may differ from the forecast readings in our data. Hence, our instrumental variable strategy can also mitigate this classical measurement error problem (Moretti and Neidell, 2011; Schlenker and Walker, 2015; Chen et al., 2018). Finally, one could also argue that the level of pollution and weather conditions in the days the exams are marked could influence the scores as the examiners could also have their productivity affected by these circumstances. However, all the tests are multiple choice and the grades are given by a machine, and thus should not be affected by air pollution or weather determinants.

4.3.2 The instrumental variable approach

After using fixed effects and controlling for weather conditions, to identify the causal relation between pollution and students' scores we partially build on Rangel and Vogl (2019)'s strategy of interacting agricultural fires with wind direction to construct our instrumental variables. Our main assumption relies upon the exogeneity of the upwind/downwind fires since it does not directly affect the children's cognitive performance, except from its effect on air pollution. As such, our instrument can be considered as a natural experiment that results in exogenous shocks to local air pollution in a specific area (Schlenker and Walker, 2015; Anderson, 2015; Deryugina et al., 2019b; Graff-Zivin et al., 2020).

The interaction of fires with wind direction is due to the concern that agricultural fires may be correlated with students' academic performance insofar as areas that use this technique can present higher returns in productivity in comparison to other locations which do not use fires to accelerate the harvesting process. Thus, in light of the possibility of agricultural fires not fulfilling the exogeneity requirement for a good instrument, we interact this variable with wind direction, which influences the spread of the fire's smokes - and as a result air pollution - but does not affect students' scores.

To deal with the endogeneity concerns above, we take advantage of the influence of the temporal and spatial variation in wind direction on the spread of the fires' smoke, and hence on air pollution levels in the nearby municipalities to construct our instruments. To this end, we count the amount of fires upwind and downwind up to 50km from the municipalities' population centroids during each of the three trimesters of pregnancy. Following Rangel and Vogl (2019), we define the wind directions using the 45 degrees.[6] In our main specification, we include the counts of upwind and downwind fires per trimester divided by the amount of days of each period. In two robustness exercises we also include the three directions, i.e., upwind, downwind and vertical wind, or sum up the counts of fires downwind with vertical as a unique category along with upwind fires.

The underpinning assumption of our instrumental variable approach is that fires located upwind from the child's municipality of birth elevate pollution more than downwind fires. As such, in this section, we present the reduced form equation to estimate the effects of agricultural fires on air pollution and the final impact of the latter on scores. The count of upwind fires is a strong predictor of local $PM_{2.5}$ levels as Table 4.2 shows. Therefore,

[6]Graff-Zivin et al. (2020) explains that there is a trade off between capturing more fires per angle (using wider angles) and more precision in determining wind direction (using narrower angles). The authors show that 45 degrees is the octant that best captures pollution dispersion and are followed by Graff-Zivin et al. (2020).

our model can be formally described as:

$$Pollutant_{mi} = \gamma upwindfires_{mi} + \zeta downwindfires_{mi} + \psi W_{mi} + \alpha X_j + \mu_m + \iota_i + \theta_t + \epsilon_{mni} \tag{4.4}$$

$$Y_{jmit} = \beta \widehat{Pollutant}_{mi} + \omega W_{mi} + \rho X_j + \nu_m + \lambda_i + \delta_t + \eta_{jmit} \tag{4.5}$$

where, in both equations, the variables are the same as in Equation 4.3, except for the outcome variable in Equation 4.4, which here is the pollutant measure linked to students j, in municipality m and time i, taking exams in t (Eq. 4.4), and the main explanatory variable in equation 4.5, which is the predicted value from the reduced form (Eq. 4.4). The coefficient of interest in equation 4.4 is γ, which informs the estimated effect of upwind fires on $PM_{2.5}$. The pollution level and weather controls correspond to the monthly average during each of the three trimesters of pregnancy, and the fire counts, to the amount of fires in each trimester. In Equation 4.5, we are interested in β as it reports the estimated causal effects of pollution on pupils' scores in the years of the examination (2011 to 2017). The other control variables, such as students' characteristics and parents' level of education, refer also to the time of the tests. We add municipalities fixed effects to control for unobserved factors that do not vary over time, for year of examinations and year and month of birth as explained before.

The literature discusses the possibility of avoidance behaviour by parents during pregnancy. For example, parents may change their behaviour in locations where there exist pollution levels alerts (Neidell, 2006, 2009; Graff-Zivin et al., 2011; Moretti and Neidell, 2011; Bharadwaj et al., 2017). Despite that, we are not aware of air pollution alerts system within our study region. As for other types of avoidance behaviour, such as closing windows, it is not possible to account for with our data.

Lastly, a limitation for studies on fetal shocks effects later in life is how to address the influence of time-varying investments parents may make in human capital to overcome fetal exposure to pollution until the exams years, which is not possible to control for in our sample (Bharadwaj et al., 2017; Almond et al., 2018). Nonetheless, other works have shown that the compensating parental responses present very small effects (Bharadwaj et al., 2013; Halla and Zweimüller, 2014; Bharadwaj et al., 2018). Therefore, if there is still some sort of unobservable investments that parents make to compensate cognitive impairments stemming from in utero pollution exposure, they will be contained in our error term and represent another negative externality from air pollution revealed as additional parental costs.[7]

4.4 Results

4.4.1 Mechanisms: Relevance of upwind fires as instrument

From Table 4.2 one can see that upwind fires are strongly causally related to the level of air pollution per trimester of pregnancy. Columns 1 to 3 refer to equations including upwind and downwind fires, while Columns 4 to 6 show results for specifications using downwind and vertical wind in a same category, and Columns 7 to 9 contain the results from the specifications which include the three categories of wind. Since we have three endogenous variables - the average $PM_{2.5}$ by trimester - we need three first stage equations. Moreover, we interact each wind direction with fires also per trimester. Therefore, there are enough instruments to support three endogenous variables within the same equation.

From the results one can see that upwind fires persistently affect pollution more than the other directions across the different forms for which we build the instruments set. This reinforces the main assumption that we rely on that local air quality is more affected

[7] Almond et al. (2018) reports a thorough review of the literature on fetal shocks.

by fires located upwind from the municipality's population centroid. In addition, and in line with Lai et al. (2018) and Graff-Zivin et al. (2020), we confirm that poor air quality is the channel through which agricultural fires affect humans' cognitive health.

4.4.2 Second Stage: Impact of air pollution on students' scores

Table 4.3 reports the results regarding the effects of $PM_{2.5}$ per three trimesters of pregnancy on students' scores. All models include flexible weather controls, and the equations estimated by Two Stage Least Squares (2SLS) include the share of fires (weighted by their probability of occurrence) per wind direction. We start by presenting our OLS estimates with Portuguese (Panel A Columns 1 and 2) and Maths (Panel B Columns 1 and 2) scores. These results show that $PM_{2.5}$ in the first and second trimesters seem to have a positive impact on Portuguese and on Maths, and no effects on the third gestational period. However, for the reasons explained above, OLS estimates are likely biased. For example, if the students are from poorer municipalities and more polluted areas, OLS may overestimate the effects of air pollution on their scores. On the other hand, if the student comes from a wealthier and polluted area, parents' compensating behaviour may not be captured by that estimator, resulting in a underestimated pollutant's coefficient Ebenstein et al. (2016). We present results using instrumental variables with and without controlling for students' characteristics (Columns 3 and 4). Comparing both columns, one notice that the magnitude of the estimates does not seem to change much by including students' controls.

We then turn to our main specifications using the counts of upwind and downwind fires as instrumental variables. One can observe that an increase of 10 $\mu g/m^3$ of $PM_{2.5}$ during the third trimester of pregnancy causes literacy reasoning to decrease by 0.21SD (Column 4 Panel A) and numeracy performance by 0.14SD (Column 4 Panel B). With regard to this, the literature shows evidence that different parts of the brain perform different functions with white matter being linked to language and verbal reasoning skills

and gray matter associated with mathematical skills (Popescu et al., 2019; Houston et al., 2019). In addition, pollution affects the white and gray matter differently, corroborating our findings suggesting that air pollution generates slightly different effects on exam performance by subject (Clougherty, 2010; Calderón-Garcidueñas et al., 2011; Erickson et al., 2020).

When we look at the first trimester point estimates, we note that the instrumental variable specification correctly dealt with the signs; i.e., contrary to the OLS likely biased positive estimates, the quasi-experimental identification now shows a negative impact although not significant. In addition, the exposure to this pollutant during the second trimester continues to spuriously suggest that pollution leads to an increase in both tests scores. These two results relating to earlier gestational periods are not as surprising as one might think since the literature points to the existence of selection bias when it comes to backwards dating of exposure to air pollution and the exact date of birth. Thus, this is a recurrent issue in several works on in utero shocks. For example, Rangel and Vogl (2019)'s findings also suffer from the same limitation.

4.4.3 Heterogeneity

In this section we explore the heterogeneous effect of the treatment effects (air pollution) for our preferred specification presented in Table 4.3 (Column 4): (i) per gender and (ii) across the distribution of scores per municipality. Through this heterogeneity analysis our goal is to unveil how different sub-populations have their cognitive performance later in life affected by the fetal exposure to poor air quality.

With regards to gender differentials, overall our findings reported in Table 4.4 suggest that boys are more negatively affected by air pollution than girls. Column 1 repeats the results of our baseline model (the same from Table 4.3) for ease of comparison to girls' and

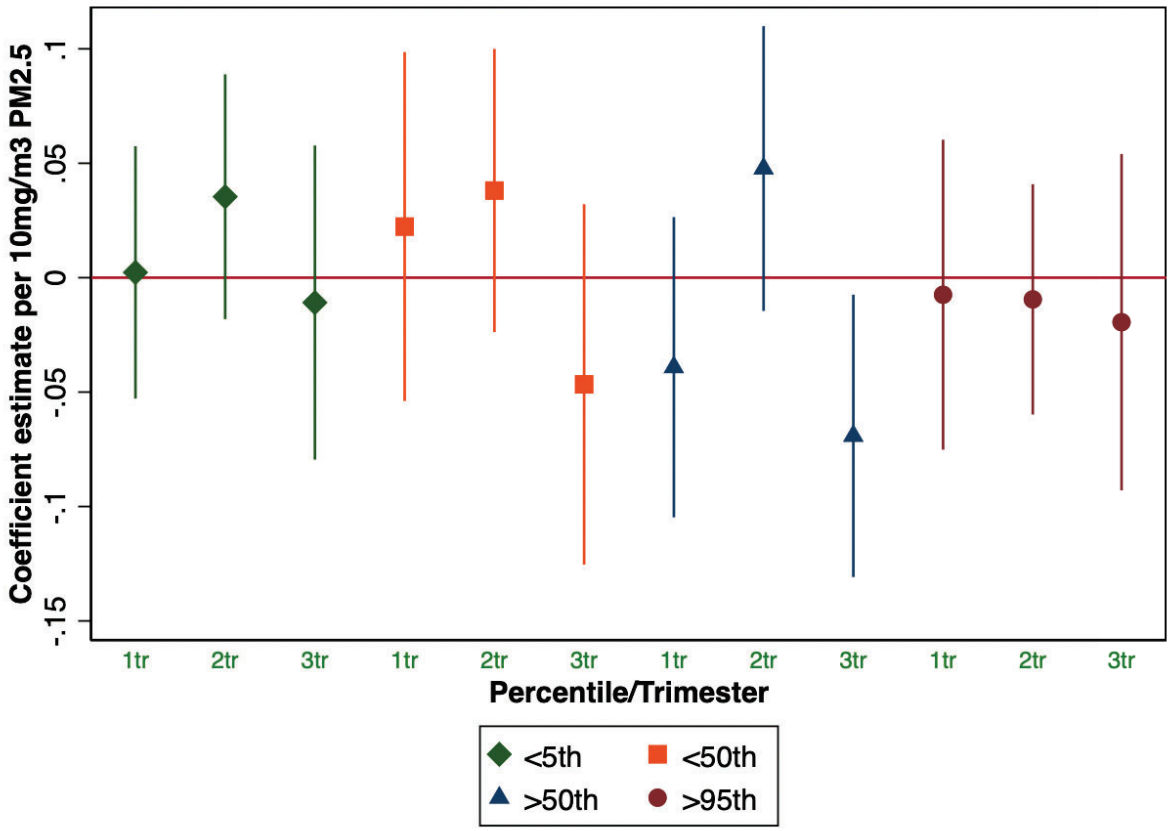
boys' performances (Columns 2 and 3) on Portuguese exams. Firstly, one can note that the pattern of signs and significance concerning the first two gestational periods remain the same, with the exposure to $PM_{2.5}$ during the first trimester showing not significant effects on cognitive performance at school, and the second trimester appearing to be positive. To put this into perspective, when it comes to the third trimester, girls seem to be less impacted by the pollutant than boys, with the latter suffering a decrease of 0.29SD in their Portuguese marks for an additional 10 $\mu g/m^3$ of $PM_{2.5}$, whereas their female counterparts have their scores dropped by 0.12SD with less significance (10%). Turning to Maths scores (Columns 4 (main specification) to 6), the results are similar, with boys' marks decreasing by 0.19SD per a rise of 10 $\mu g/m^3$ of $PM_{2.5}$, and girls' numeracy reasoning not significantly affected by pollution. This finding is consistent with previous studies including Ebenstein et al. (2016) and the results in Chapter 3, which suggest that males suffer a larger drop in their tests' scores when exposed to poor air quality during the examinations.

Although the role of gender in the effects of air pollution on respiratory health remains unclear, some epidemiology studies point out that males are more affected than females. For example, Abbey et al. (1998) link PM_{10} exposure to reduced lung function among males but not females, Galizia and Kinney (1999) find similar results regarding the exposure to O_3, and Chen et al. (2017b) identify that contemporaneous and cumulative exposure to air pollution has a stronger negative effect for men than for women. Similarly, Ebenstein et al. (2016) find the effect of pollution on exam performance to differ by gender and cite the increased incidence of asthma amongst boys as a possible cause.

As per the second heterogeneity analysis, we aim to detect whether strong or middling students are differently affected by poor air quality. To this end we split our final sample into subsets by low and high scores within each municipality with results for Portuguese and Maths scores as dependent variables depicted in Figures 4.1 and 4.2. One

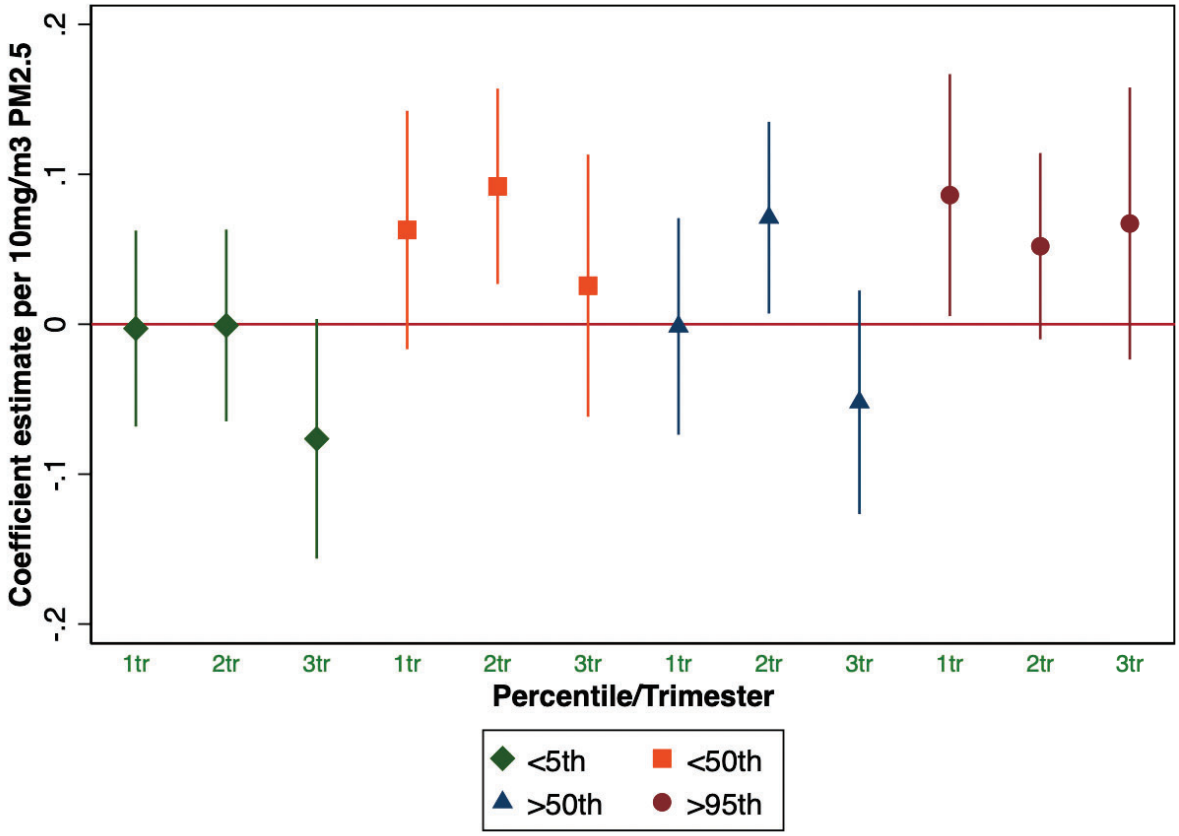
can note that weaker students (5^{th} percentile) are more sensitive to exposure during the third trimester of pregnancy than the others in the scores distributions for both subjects. Nonetheless, students above the median also have their marks diminished as a result of fetal exposure to air pollution during the third period of pregnancy. Noteworthy is the fact that top students within the 95^{th} percentile seem also to have their Maths' scores reduced. Both results are in line with Ebenstein et al. (2016) who also show evidence of weaker students' poorer performance as a result of bad air quality during the university examinations, and Graff-Zivin et al. (2020) who show that students belonging to the top of the distribution are more affected by bad air quality.

Figure 4.1: Heterogeneity in Particulate Matter's Impact on Students' Scores Across Portuguese Scores' Distribution



Note: Plot of coefficients from four regressions analysing heterogeneity across the Portuguese score distribution. The x axis presents the coefficient of $PM_{2.5}$ per each gestational trimester. Each group of symbols represents the percentile explored as in the legend. The details on the specifications are the same as in table 4.3.

Figure 4.2: Heterogeneity in Particulate Matter's Impact on Students' Scores Across Maths Scores' Distribution



Note: Plot of coefficients from four regressions analysing heterogeneity across the Maths score distribution. The x axis presents the coefficient of $PM_{2.5}$ per each gestational trimester. Each group of symbols represents the percentile explored as in the legend. The details on the specifications are the same as in table 4.3.

4.4.4 School Performance and Wages

The literature on the returns of education has extensively showed that better cognitive test scores at school are positively related to individual higher earnings in the future (Murnane et al., 1995; Murphy and Peltzman, 2004; Blau and Kahn, 2005; Chetty et al., 2011; Currie et al., 2012; Curi and Menezes-Filho, 2015; Bharadwaj et al., 2017), to lower probability of drop-outs (Rivkin, 1995; Hanushek, 1996); and to higher economic growth (Hanushek and Kimko, 2000).

Curi and Menezes-Filho (2015) shows evidence of a 5% reduction of salary per 10% reduction of scores. Therefore, using our estimated cognitive performance effects and estimates on the returns to exams scores from Brazil (Curi and Menezes-Filho, 2015), we

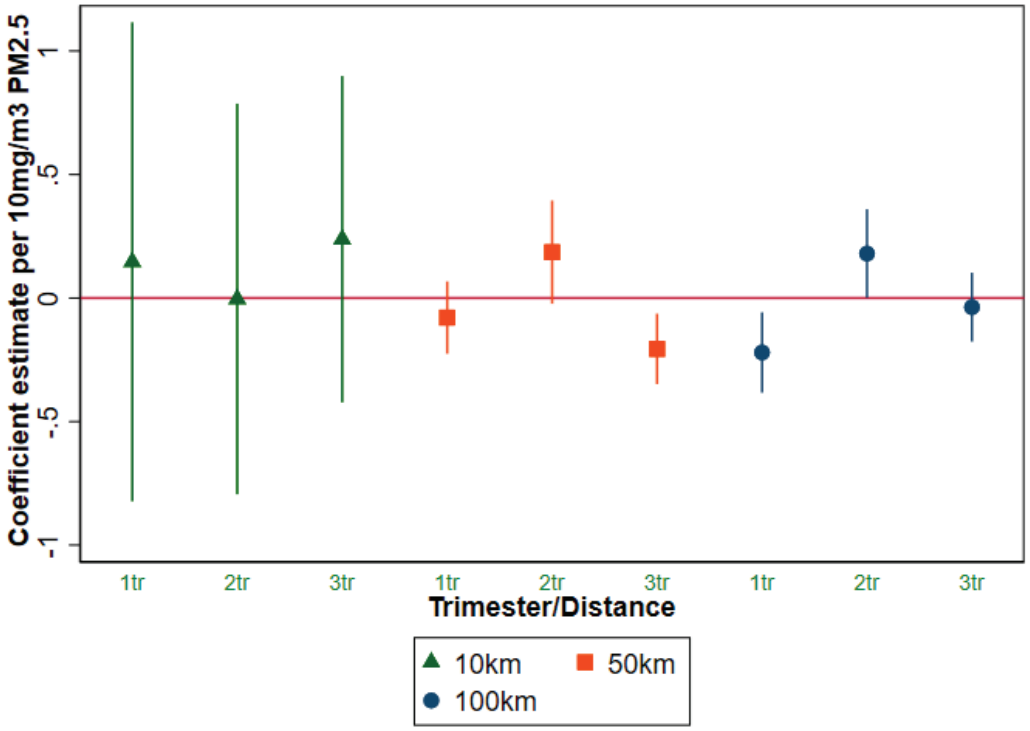
present a back-of-the-envelope calculation suggesting that a drop of $PM_{2.5}$ by 10% (*ceteris paribus*) would rise students' scores by 4.6%; and consequently elevate future individuals' earnings by 2.3%. Our calculations are based on Portuguese estimates. Overall, what we found is that being exposed to $10\mu g/m^3$ extra of $PM_{2.5}$ during the third trimester of pregnancy leads to a decrease in scores in the 5^{th} grade by 4.6% ((0.21 X 47.8)/219 = 4.6), where |0.21| is the coefficient from main estimation from Table 4.3, 47.8 and 219 are the SD and the average of Portuguese scores from Table 4.1). Bharadwaj et al. (2017) also found similar results when investigating fetal exposure to air pollution and students' scores in the fourth grade in Chile. The authors use the study on the returns to test scores in the United States by Blau and Kahn (2005) to provide their back-of-envelope calculation for Chilean students suggesting that, *ceteris paribus*, a reduction of 50% of Carbon Monoxide (CO) levels propels additional 1,000 US dollars in lifetime earnings per infant belonging to the cleaner cohort.

4.5 Sensitivity Analysis

To assess the internal validity of our research design and results we carry out a set of robustness tests. In the first exercise, we vary the distance between the fires' location and the municipalities' population centroid, using 10, 50 and 100 km (50km used in the main results). From Figures 4.3 and 4.4, one can observe that the effect of air pollution diminishes when the distance from the fires increases. This suggests that the pollutants' levels decay insofar as they travel, results consistent with Lai et al. (2018)'s findings.

Differently from the main specification in which the instruments used were the count of fires upwind and downwind, our second robustness exercise adds vertical fires as as instrument . The third category accounts for fires that are not upwind or downwind from the municipality's population centroid. Table 4.5 Columns 1-3 present results for the

Figure 4.3: Cognitive (Portuguese scores) impact of $PM_{2.5}$ per distance from the fires points to the population's municipality centroid

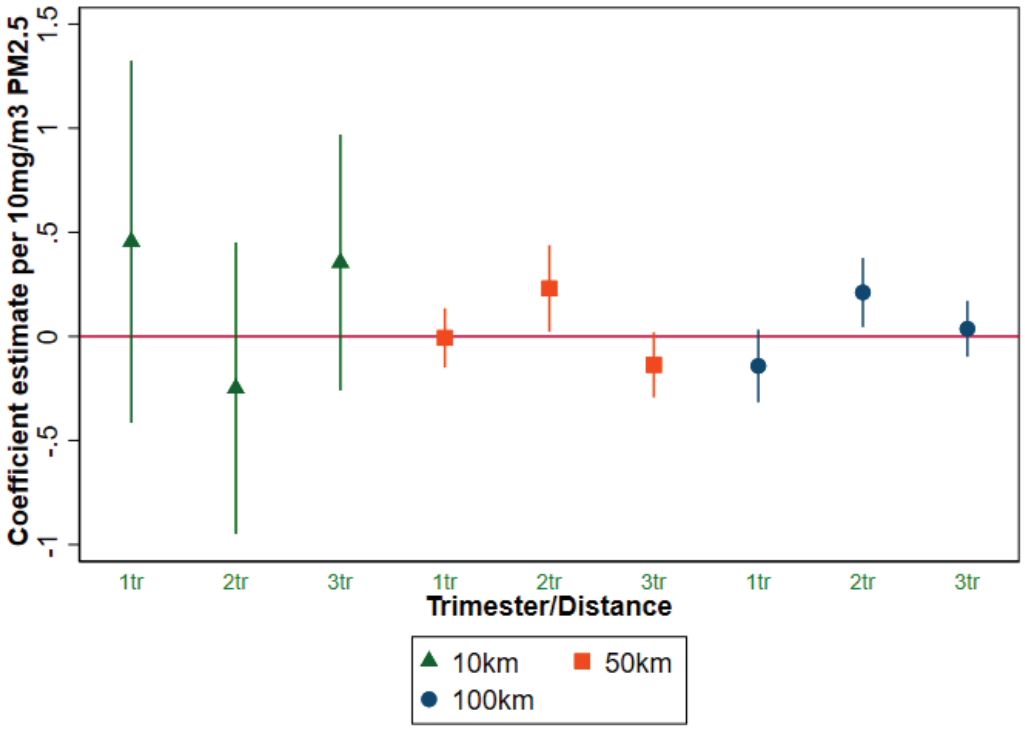


Note: This figure reports the cognitive (Portuguese scores) impact of air pollution within different distances. 10km is the baseline model. All the specifications follow the main model from table 4.3. We note an increase of standard errors suggesting that the further the stations the less precise and significant the estimates mainly for the third trimester.

outcome of Portuguese scores and Columns 4-6 for Maths. Columns 2 and 5 report the estimates for models with aggregated downwind and vertical fires in the same category as reported, with coefficients qualitatively unchanged when compared to the main specifications (Columns 1 and 4). In addition, in Columns 3 and 6, we utilise specifications with the three types of fires as instruments; i.e., upwind, downwind and vertical fires in the same equation. This latter test also shows that the main results (Column 1 and 4) remain qualitatively unchanged in terms of magnitude, signs, significance and the strength of instruments.

In our third set of robustness checks we slightly change the way we build our fire counts. In our baseline model, we utilise the confidence-weighted fire measure as the NASA's data set attributes a confidence index to each pixel-day as a measure of the prob-

Figure 4.4: Cognitive (Maths scores) impact of $PM_{2.5}$ per distance from the fires points to the population's municipality centroid



Note: This figure reports the cognitive (Maths scores) impact of air pollution within different distances. 10km is the baseline model. All the specifications follow the main model from table 4.3. We note an increase of standard errors suggesting that the further the stations the less precise and significant the estimates mainly for the third trimester.

ability that a fire occurred, ranging from 0% to 100%, based on weather conditions and vegetation at the exact moment the satellite captured the potential fire (Setzer and Sismanoglu, 2007; Rangel and Vogl, 2019). From Table 4.6 (Column 4 to 6), one can observe that the magnitude of the coefficients of the non-confidence weighted fires are somewhat smaller than the probability-weighted fires used in the main specification, with the First Stage F test still indicating strong instruments. Additionally, Table 4.7 suggests a slight shrinkage of the magnitude of air pollution's coefficients predicted utilising non-weighted fires (Columns 3 and 4).

To demonstrate that our baseline estimates with binned weather controls are not sensitive to different forms of meteorological controlling, we include a fourth robustness analysis. Table 4.8 presents our main specification results without accounting for weather

and with less flexible controls, i.e., with linear, squared, and interactions controls). The results remain generally the same, indicating that our main estimates are not driven by unobserved weather conditions correlated with wind and students' performance, but solely by the effects of air pollution.

In a fifth robustness exercise, Table 4.9 presents results of our estimations with SO_2 as co-pollutant ($PM_{2.5}$ and SO_2 present a pairwise correlation of 27.3% significant at 5% level). One can notice that the First Stage F statistics for both pollutants drops well below the threshold of 10, indicating that our fires instruments are not relevant to predict the latter pollutant. It is somehow a good sign suggesting that our baseline results using only $PM_{2.5}$ are not driven by other pollutants; that is, our instrumental variables approach helps us to disentangle the effect of $PM_{2.5}$ from other components of air pollution.

4.6 Conclusion

In this chapter, we examine the relationship between fetal exposure to air pollution from agricultural fires and cognitive performance on fifth grade exams in Brazil. To this end we merge data from the Brazilian Ministry of Education; NASA's satellite remote sensing data with active fires detection and air pollution data and atmospheric conditions gridded data from the Japanese Reanalysis archive, JRA-55. We use a set of controls for parents' level of education and children's characteristics to control for observable effects, as well as for municipalities unobserved fixed effects. Despite the inclusion of these control variables, some unobserved time-varying features can cause endogeneity in our analysis due to omission and classical measurement error inherent to this type of air pollution study. Thus, to address this problem we use an instrumental variable approach that we consider capable of creating exogenous shocks to our air pollution measure, $PM_{2.5}$.

Our empirical strategy, which relies upon changes in wind direction and fires' location as predictors of air pollution, yields credible evidence that the exposure to poor air quality during the gestational period causes a reduction in cognitive ability of students. More precisely, we argue that students living in municipalities located upwind from the fires present lower academic performance than their downwind dwellers counterparts. To put our findings into perspective, we demonstrate that students exposed to additional $10\mu g/m^3$ of $PM_{2.5}$ during the third trimester of pregnancy suffer a reduction of 4.6% in their Portuguese scores and 2.8% in Maths. On top of that, when we study the response from different sub-populations within our sample, we find that boys are more affected than girls. For example, males' literacy skills drop by 0.29SD, whereas females' Portuguese scores are reduced by 0.12SD per extra $10\mu g/m^3$ of $PM_{2.5}$ during the third gestational period. When it comes to numerical ability, boys' Maths scores fall by 0.19SD per $10\mu g/m^3$ of that pollutant, while girls' seem not to be significantly affected. These findings are consistent with previous studies on the impact of poor air quality and cognitive impairment including Jayachandran (2009), Sanders (2012), Ebenstein et al. (2016) and the third chapter of this thesis. Although the role of gender in the effects of air pollution on respiratory health remains unclear, some epidemiology studies point out that males are more affected than females. For example, Abbey et al. (1998) link PM_{10} exposure to reduced lung function among males but not females; and Galizia and Kinney (1999) find similar results regarding the exposure to O_3. Similarly, Ebenstein et al. (2016) find the effect of pollution on exam performance to differ by gender and cite the increased incidence of asthma amongst boys as a possible cause. Another interesting result emerged when exploring the scores' distribution. We find that very weak students (5^{th} percentile) and very top students (95^{th} percentile) are more sensitive to exposure during the third trimester of pregnancy than other pupils for both subjects. These results are in line with previous works, for example, Ebenstein et al. (2016) find evidence of students from the bottom of the distribution being more negatively impacted by bad air quality during the examinations, whereas, in Graff-Zivin et al. (2020)'s study, stronger pupils have their

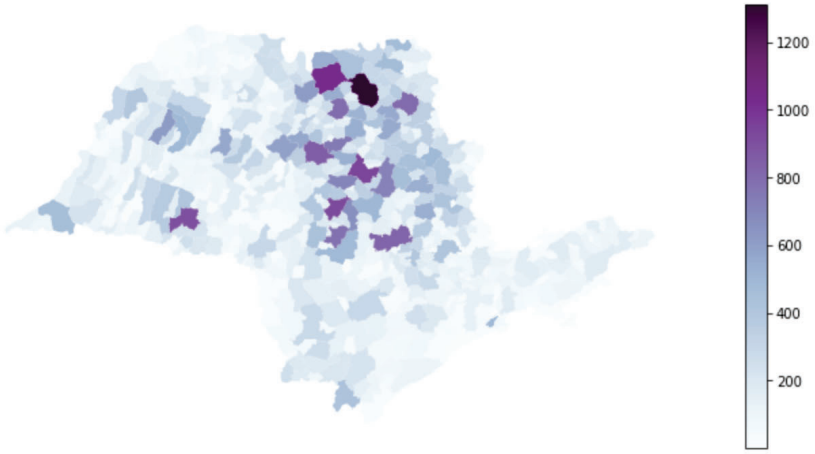
scores worsened by the level of pollution in the day of the tests.

To check the internal validity of our findings we perform a series of tests using different sets of instruments and distances between fires and municipality's population centroid, equations with co-pollutant SO_2 and falsification tests to which our results are robust and consistently suggest the harmful effect of fetal exposure to air pollution and cognitive health. In addition, we perform back-of-the-envelope calculations and show that a reduction of 10% of $PM_{2.5}$ during the third trimester of pregnancy has the potential to increase later life wages by 2.3%. These findings shed light on the important issue of wealth accumulation mainly in developing countries with sharper levels of inequality. Suffice it to say that the less well-off, who live in areas with low environmental standards, end up suffering more from the negative impacts of fetal exposure to air pollution on their cognitive performance, and hence deepening the wage inequality in Brazil.

All in all, our results seem to add plausible evidence suggesting that cognitive performance may be hindered by in utero exposure to bad air quality. As such, this chapter's findings may be useful to cast light on the imminent necessity of withdrawing traditional agricultural burning techniques, which, in fact, is due to occur soon. In fact, the government of the state of São Paulo has signed the Green Ethanol Protocol in 2017 due to start its validity in 2021 (SIMA, 2017) aimed at boosting the mechanisation of the harvesting phase as well as restoring the riparian areas. Although the fires have not started to decline yet, they seem to have plateaued (Aguiar et al., 2011), which is a positive sign towards the reduction of air pollution and, as demonstrated in our findings, it may entail long-run gains in economic growth through the human capital channel. In the short term, future research can be conducted after the implementation of this policy to assess its impacts on birth outcomes.

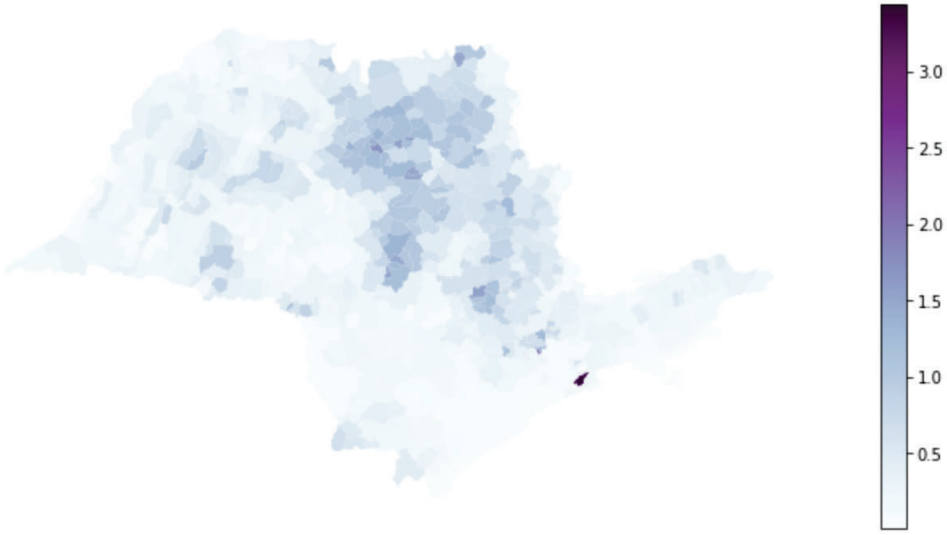
Appendix

Figure C1: Total of agricultural fires per municipality from 2000 to 2008



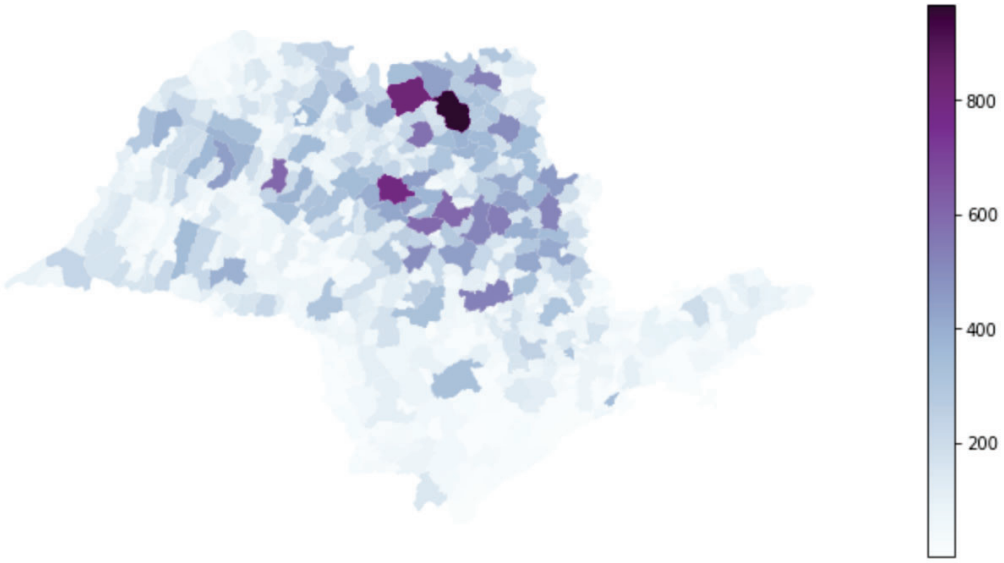
Note: Total amount of fires per municipality in the state of São Paulo during the birth years of our sample cohorts.

Figure C2: Total of agricultural fires per municipality size from 2000 to 2008



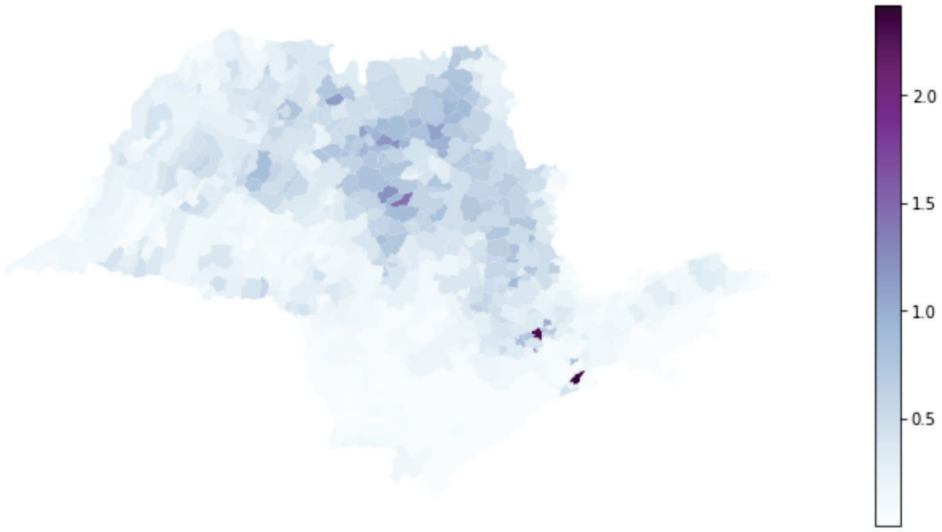
Note: Total amount of fires per square kilometers within each municipality of the state of São Paulo during the birth years of our sample cohorts.

Figure C3: Total of agricultural fires per municipality from 2009 to 2017



Note: Total amount of fires per municipality in the state of São Paulo in our samples' exams period.

Figure C4: Total of agricultural fires per municipality size from 2009 to 2017



Note: Total amount of fires per square kilometers within each municipality of the state of São Paulo in our samples' exams period.

Table 4.2: Instrument for Pollution: Count of fires per Wind Direction – First Stage Results)

Fires counts by wind direction	(1) PM_{25}_1tr	(2) PM_{25}_2tr	(3) PM_{25}_3tr	(4) PM_{25}_1tr	(5) PM_{25}_2tr	(6) PM_{25}_3tr	(7) PM_{25}_1tr	(8) PM_{25}_2tr	(9) PM_{25}_3tr
Upwind_1tr	0.00	-0.05**	-0.06**	-0.10	-0.01	-0.04	-0.09	-0.03	-0.02
	(0.05)	(0.02)	(0.03)	(0.06)	(0.02)	(0.05)	(0.06)	(0.02)	(0.06)
Downwind_1tr	0.31***	-0.14***	0.04**				0.20***	-0.12**	0.10**
	(0.07)	(0.04)	(0.02)				(0.06)	(0.06)	(0.05)
Upwind_2tr	0.05**	0.04	-0.06***	0.04*	-0.06	-0.04**	0.06***	-0.06	-0.04
	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.03)
Downwind_2tr	0.07**	0.25***	-0.04				0.06**	0.11***	-0.01
	(0.03)	(0.04)	(0.03)				(0.03)	(0.03)	(0.03)
Upwind_3tr	-0.10***	0.11***	0.06**	-0.08**	0.10***	-0.02	-0.06**	0.11***	-0.01
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
Downwind_3trsh	0.02	0.04*	0.22***				0.07*	0.01	0.15***
	(0.02)	(0.02)	(0.03)				(0.04)	(0.02)	(0.03)
Down+Vertical wind_1tr				0.15***	-0.06***	0.00			
				(0.03)	(0.01)	(0.02)			
Down+Vertical wind_2tr				0.02**	0.13***	-0.02**			
				(0.01)	(0.02)	(0.01)			
Down+Vertical wind_3tr				-0.00	0.01	0.11***			
				(0.01)	(0.01)	(0.01)			
Vertical wind_1tr							0.12***	-0.02	-0.05
							(0.04)	(0.02)	(0.05)
Vertical wind_2tr							-0.00	0.14***	-0.03
							(0.01)	(0.04)	(0.02)
Vertical wind_3tr							-0.05*	0.01	0.08***
							(0.03)	(0.02)	(0.02)
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.43	0.36	0.32	0.44	0.37	0.32	0.44	0.37	0.33

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured in 10 $\mu g/m^3$. Up, down and vertical wind represent the share of fires (weighted by confidence) per trimester of pregnancy by wind direction. Columns 1 to 3 present the first stage for the main specification with instruments upwind and downwind fires. Columns 4 to 6 present models using two winds categories: upwind; downwind + vertical wind). Columns 7 to 9 refer to models using three categories of wind: upwind, downwind and vertical wind fires confidence weighted.

Table 4.3: The Effects of $PM_{2.5}$ on Students' Scores - OLS and IV (Main Specification with flexible weather controls)

	Without Instruments		With Instruments	
Variables	(1)	(2)	(3)	(4)
	Panel A: Dep. Var. Portuguese Score (SD)			
$PM_{2.5}$_1tr	0.04**	0.03**	-0.08	-0.08
	(0.01)	(0.01)	(0.09)	(0.09)
$PM_{2.5}$_2tr	0.17***	0.15***	0.21**	0.18**
	(0.03)	(0.03)	(0.09)	(0.08)
$PM_{2.5}$_3tr	-0.03	-0.02	-0.24***	-0.21***
	(0.05)	(0.04)	(0.07)	(0.07)
Students' Controls	N	Y	N	Y
First stage F-test			14.9	14.9
Observations	1,190,717	1,187,660	1,190,717	1,187,660
R-squared	0.04	0.09	0.00	0.05
	Without Instruments		With Instruments	
Variables	(1)	(2)	(3)	(4)
	Panel B: Dep. Var. Mathematics Score (SD)			
$PM_{2.5}$_1tr	0.03	0.03*	-0.01	-0.01
	(0.02)	(0.02)	(0.08)	(0.08)
$PM_{2.5}$_2tr	0.16***	0.15***	0.24***	0.23***
	(0.02)	(0.02)	(0.09)	(0.08)
$PM_{2.5}$_3tr	-0.03	-0.03	-0.16**	-0.14*
	(0.04)	(0.04)	(0.08)	(0.08)
Students' Controls	N	Y	N	Y
First stage F-test			14.9	14.9
Observations	1,190,717	1,187,660	1,190,717	1,187,660
R-squared	0.06	0.10	0.00	0.04

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured in 10 $\mu g/m^3$. Instruments used upwind and downwind fires confidence weighted.

Table 4.4: Heterogeneity in Particulate Matter's Impact on Students' Scores per Gender

	Portuguese Scores (SD)			Mathematics Scores (SD)		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	All	Girls	Boys	All	Girls	Boys
PM_{25}_1tr	-0.08	-0.16*	-0.03	-0.01	-0.07	0.03
	(0.09)	(0.09)	(0.10)	(0.08)	(0.09)	(0.09)
PM_{25}_2tr	0.18**	0.09	0.25***	0.23***	0.14	0.29***
	(0.08)	(0.10)	(0.07)	(0.08)	(0.09)	(0.08)
PM_{25}_3tr	-0.21***	-0.12*	-0.29***	-0.14*	-0.05	-0.19**
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
First stage F-test	14.9	13.9	14.4	14.9	13.9	14.4
Observations	1,187,660	543,030	644,630	1,187,660	543,030	644,630
R-squared	0.05	0.05	0.04	0.04	0.04	0.04

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Instruments used upwind and downwind fires confidence weighted.

Table 4.5: The Effects of $PM_{2.5}$ on Students' Scores (using IV with different wind directions specifications)

	Portuguese Score (SD)			Mathematics Score (SD)		
	Main Model	Two Winds	Three Winds	Main Model	Two Winds	Three Winds
Variables	(1)	(2)	(3)	(4)	(5)	(6)
$PM_{2.5}_1tr$	-0.08	-0.10	-0.08	-0.01	-0.03	-0.01
	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)
$PM_{2.5}_2tr$	0.18**	0.15***	0.16***	0.23***	0.19***	0.20***
	(0.08)	(0.06)	(0.06)	(0.08)	(0.05)	(0.05)
$PM_{2.5}_3tr$	-0.21***	-0.23***	-0.22***	-0.14*	-0.15*	-0.14*
	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)
First stage F-test	14.9	25	16.1	14.9	25	16.1
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.05	0.05	0.05	0.04	0.04	0.04

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured in 10 $\mu g/m^3$. Column 1 presents the main specification with instruments upwind and downwind fires. Columns 2 and 5 present models using two winds categories: upwind; downwind + vertical wind). Columns 4 and 6 refer to models using three categories of wind: upwind, downwind and vertical wind fires confidence weighted.

Table 4.6: First Stage Results: Instruments for Pollution using non-confidence weighted fires

	Confidence Weighted Fires			Non Confidence Weighted Fires		
Fires counts by	(1)	(2)	(3)	(4)	(5)	(6)
wind direction	PM_{25}_1tr	PM_{25}_2tr	PM_{25}_3tr	PM_{25}_1tr	PM_{25}_2tr	PM_{25}_3tr
Upwind_1tr	0.00	-0.05**	-0.06**			
	(0.05)	(0.02)	(0.03)			
Downwind_1tr	0.31***	-0.14***	0.04**			
	(0.07)	(0.04)	(0.02)			
Upwind_2tr	0.05**	0.04	-0.06***			
	(0.02)	(0.03)	(0.02)			
Downwind_2tr	0.07**	0.25***	-0.04			
	(0.03)	(0.04)	(0.03)			
Upwind_3tr	-0.10***	0.11***	0.06**			
	(0.03)	(0.02)	(0.02)			
Downwind_3tr	0.02	0.04*	0.22***			
	(0.02)	(0.02)	(0.03)			
Upwind_1tr				-0.03	-0.03**	-0.05***
				(0.04)	(0.01)	(0.02)
Downwind_1tr				0.24***	-0.11***	0.04***
				(0.05)	(0.03)	(0.01)
Upwind_2tr				0.02	0.00	-0.04***
				(0.02)	(0.02)	(0.01)
Downwind_2tr				0.07***	0.18***	-0.04*
				(0.03)	(0.02)	(0.02)
Upwind_3tr				-0.07***	0.08***	0.02
				(0.02)	(0.01)	(0.02)
Downwind_3tr				0.01	0.04***	0.18***
				(0.02)	(0.01)	(0.02)
First stage F-test	14.9	14.9	14.9	14.8	14.8	14.8
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660

Notes: Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Columns 1 to 3 report results using upwind and downwind fires weighted by the probability of occurrence; Columns 4 to 6 report results using upwind and downwind fires non-weighted by the probability of occurrence.

Table 4.7: Particulate Matter's Impact on Students' Scores using fires non-confidence weighted

	Portuguese Scores (SD)		Mathematics Scores (SD)	
	(1)	(2)	(3)	(4)
	Confidence	Non-confidence	Confidence	Non-confidence
Variables	Weighted Fires	Weighted Fires	Weighted Fires	Weighted Fires
PM_{25}_1tr	-0.08	-0.09	-0.01	-0.03
	(0.09)	(0.08)	(0.08)	(0.07)
PM_{25}_2tr	0.18**	0.21**	0.23***	0.25***
	(0.08)	(0.09)	(0.08)	(0.08)
PM_{25}_3tr	-0.21***	-0.21***	-0.14*	-0.15*
	(0.07)	(0.07)	(0.08)	(0.08)
First stage F-test	14.9	14.8	14.9	14.8
Observations	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.05	0.05	0.04	0.04

Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Instruments used upwind and downwind fires non-confidence weighted.

Table 4.8: The Effects of $PM_{2.5}$ on Students' Scores (SD) (without and with linear weather controls)

	Without IV		With IV		Without IV		With IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Portuguese Scores (SD)				Mathematics Scores (SD)			
PM_{25}_1tr	-0.01	-0.10	0.03*	-0.06	-0.02	-0.05	0.02	0.01
	(0.01)	(0.06)	(0.01)	(0.09)	(0.01)	(0.06)	(0.02)	(0.08)
PM_{25}_2tr	0.12***	0.13	0.15***	0.22**	0.11***	0.17**	0.14***	0.26***
	(0.02)	(0.08)	(0.02)	(0.09)	(0.02)	(0.08)	(0.02)	(0.09)
PM_{25}_3tr	-0.02	-0.07	-0.02	-0.19***	-0.02	0.01	-0.02	-0.12
	(0.03)	(0.08)	(0.04)	(0.07)	(0.03)	(0.08)	(0.04)	(0.08)
Weather Controls	N	N	Y	Y	N	N	Y	Y
First stage F-test			29.2	16			29.2	16
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.09	0.05	0.09	0.05	0.09	0.04	0.10	0.04

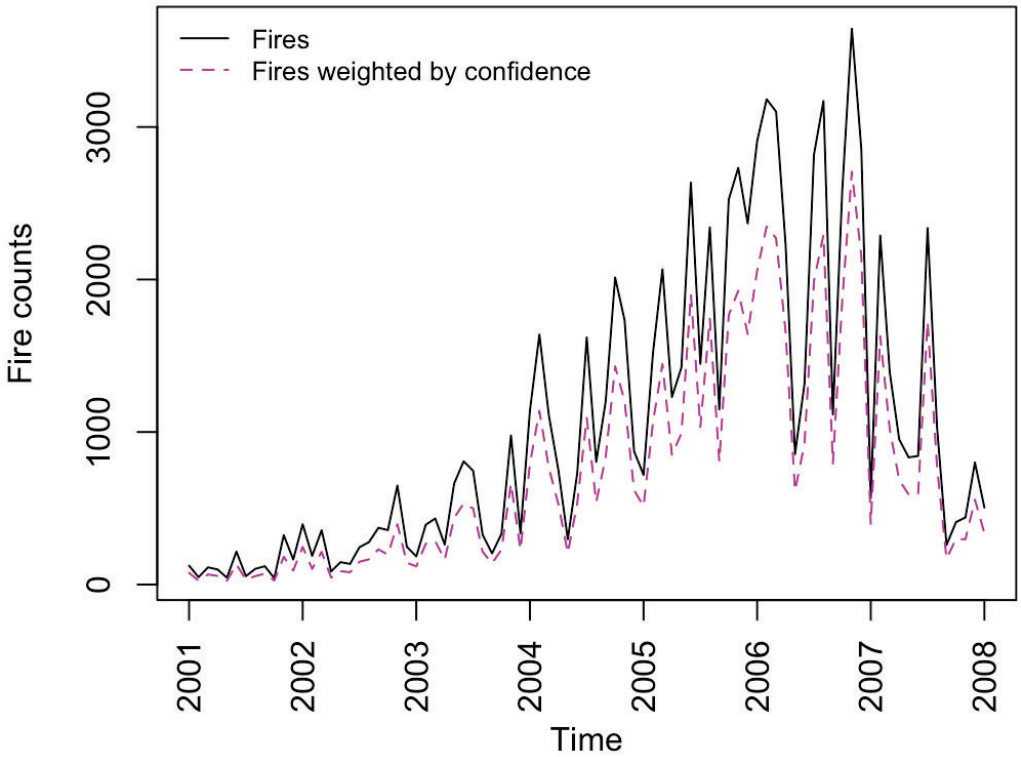
Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature linear, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr and PM_{25}_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Instruments used upwind and downwind fires confidence weighted.

Table 4.9: The Effects of $PM_{2.5}$ on Students' Scores (SD) (with co-pollutant)

	Without IV		With IV		Without IV		With IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Portuguese Scores (SD)				Mathematics Scores (SD)			
PM_{25}_1tr	0.03**	0.04***	-0.08	-0.08	0.03*	0.04***	-0.01	-0.01
	(0.01)	(0.02)	(0.09)	(0.08)	(0.02)	(0.02)	(0.08)	(0.08)
PM_{25}_2tr	0.15***	0.18***	0.18**	0.19**	0.15***	0.19***	0.23***	0.24***
	(0.03)	(0.03)	(0.08)	(0.08)	(0.02)	(0.03)	(0.08)	(0.08)
PM_{25}_3tr	-0.02	-0.01	-0.21***	-0.21***	-0.03	-0.01	-0.14*	-0.15*
	(0.04)	(0.04)	(0.07)	(0.07)	(0.04)	(0.03)	(0.08)	(0.08)
SO_2_1tr		0.02		0.00		0.00		0.00
		(0.04)		(0.00)		(0.04)		(0.00)
SO_2_2tr		-0.12***		0.00		-0.15***		0.00
		(0.04)		(0.00)		(0.05)		(0.00)
SO_2_3tr		-0.11*		0.00		-0.11		0.00
		(0.06)		(0.00)		(0.07)		(0.00)
First stage F-test			14.9	0.35			14.9	0.35
Observations	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660	1,187,660
R-squared	0.09	0.09	0.05	0.05	0.10	0.10	0.04	0.04

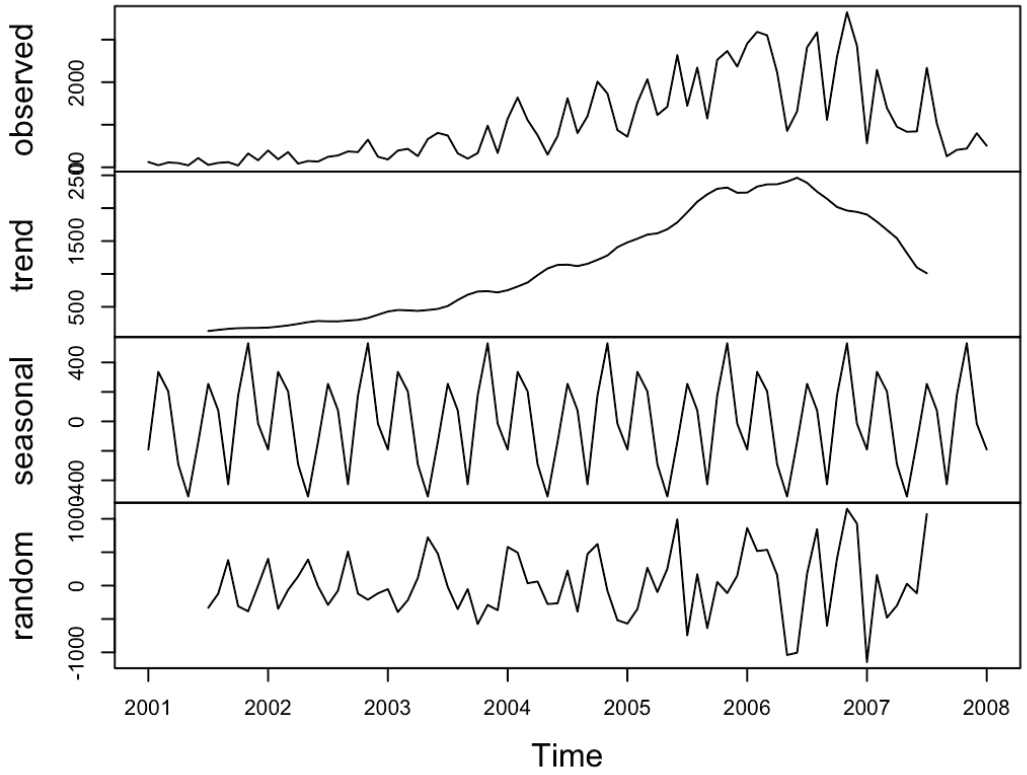
Notes: Robust standard errors in parentheses clustered by municipality centroid. *** p<0.01, ** p<0.05, * p<0.1. First stage Kleibergen-Paap rk Wald F statistic. All models control for municipality FE, day and year of exam, month and year of birth, and students' controls. Estimates for controls not shown: temperature in bins, rainfall and its square, relative humidity and its square, interactions of these weather variables, and wind speed. PM_{25}_1tr, PM_{25}_2tr, PM_{25}_3tr, SO_2_1tr, SO_2_2tr and SO_2_3tr account for their linear form for trimester 1, 2 and 3 of pregnancy measured 10 $\mu g/m^3$. Instruments used upwind and downwind fires confidence weighted.

Figure C5: Monthly average of counts of fires from 2001-2008



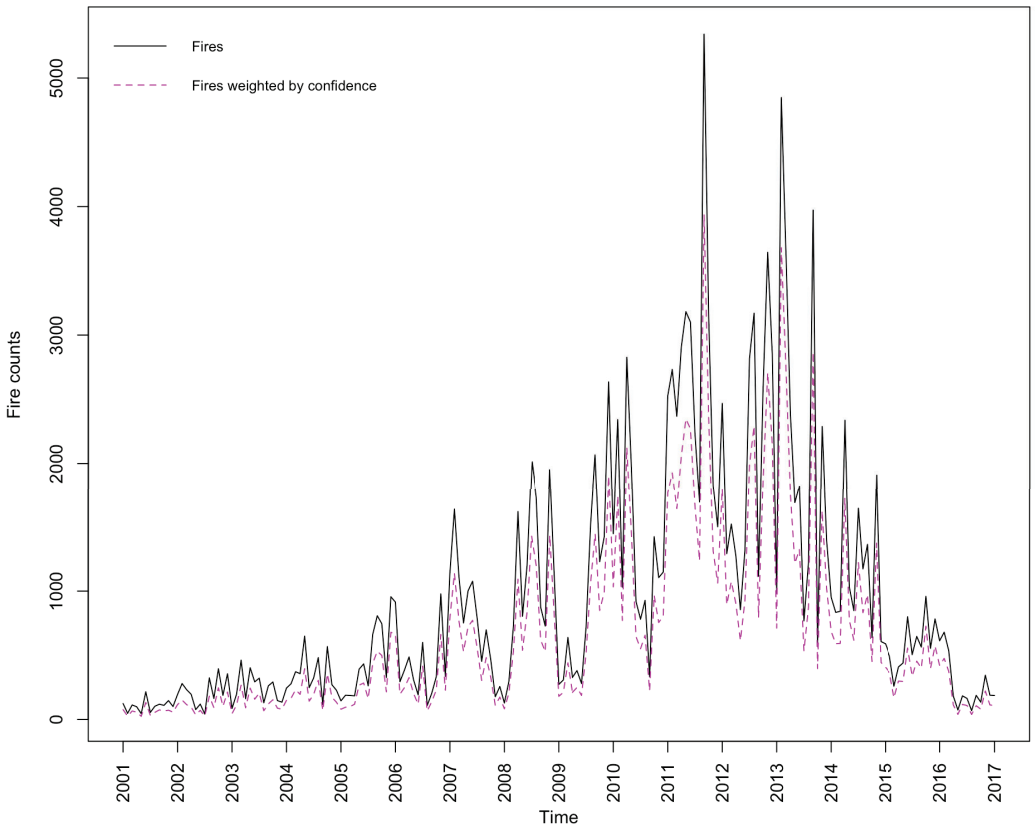
Note: The graph shows the monthly average of counts of fires and fires confidence-weighted from 2001-2008 in São Paulo state. We can see a that the peak of fires occur during the winter months.

Figure C6: Decomposition Additive Time Series of monthly average counts of fires from 2001-2008



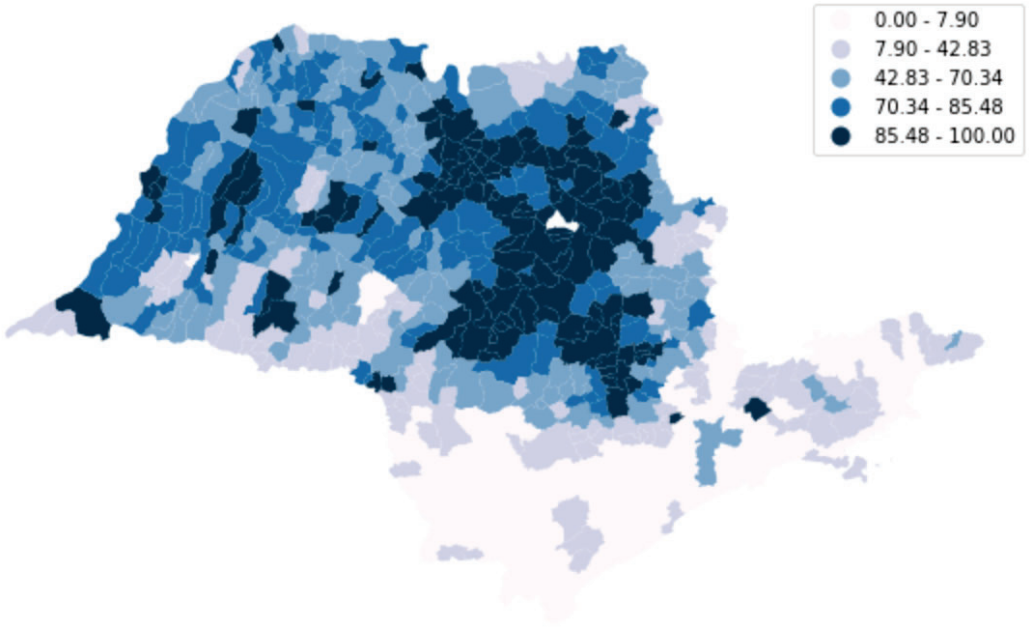
Note: The graphs shows Decomposition Additive Time Series of monthly average counts of fires from 2001-2008 in São Paulo state. We can see a strong seasonality with the peak of fires occur during the winter months and a not clear time trend.

Figure C7: Monthly average of counts of fires from 2001-2017



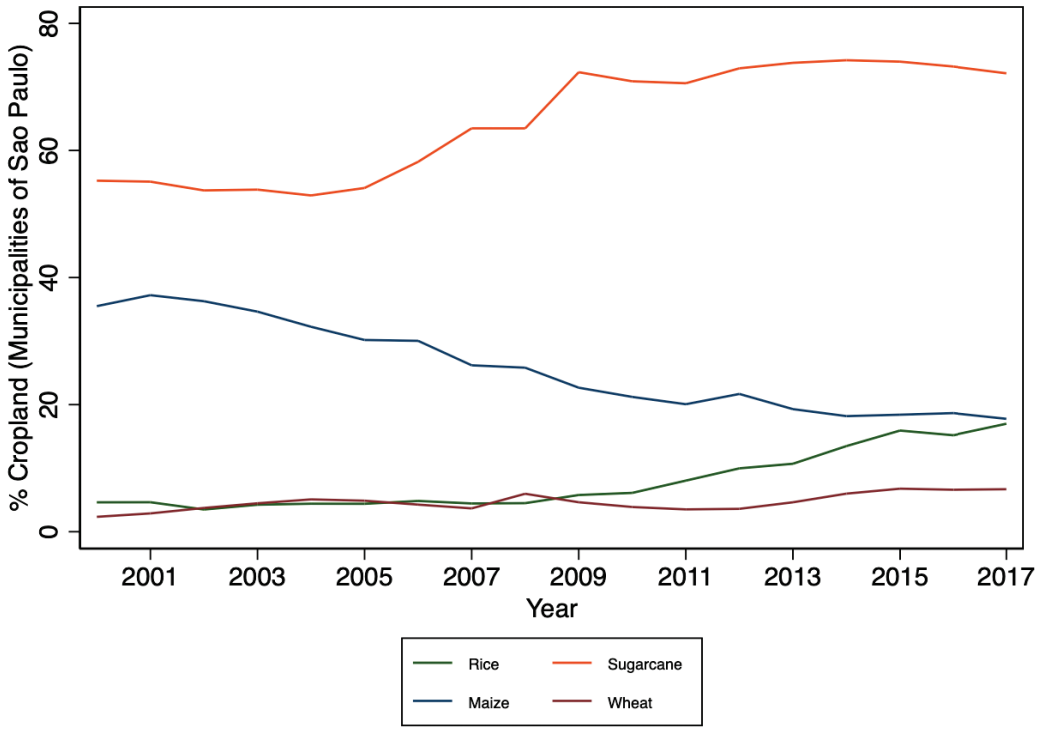
Note: The graph shows the monthly average of counts of fires and fires confidence-weighted from 2001-2017 in São Paulo state. We can see a that the peak of fires occur during the winter months.

Figure C8: Share of sugarcane crop per municipality



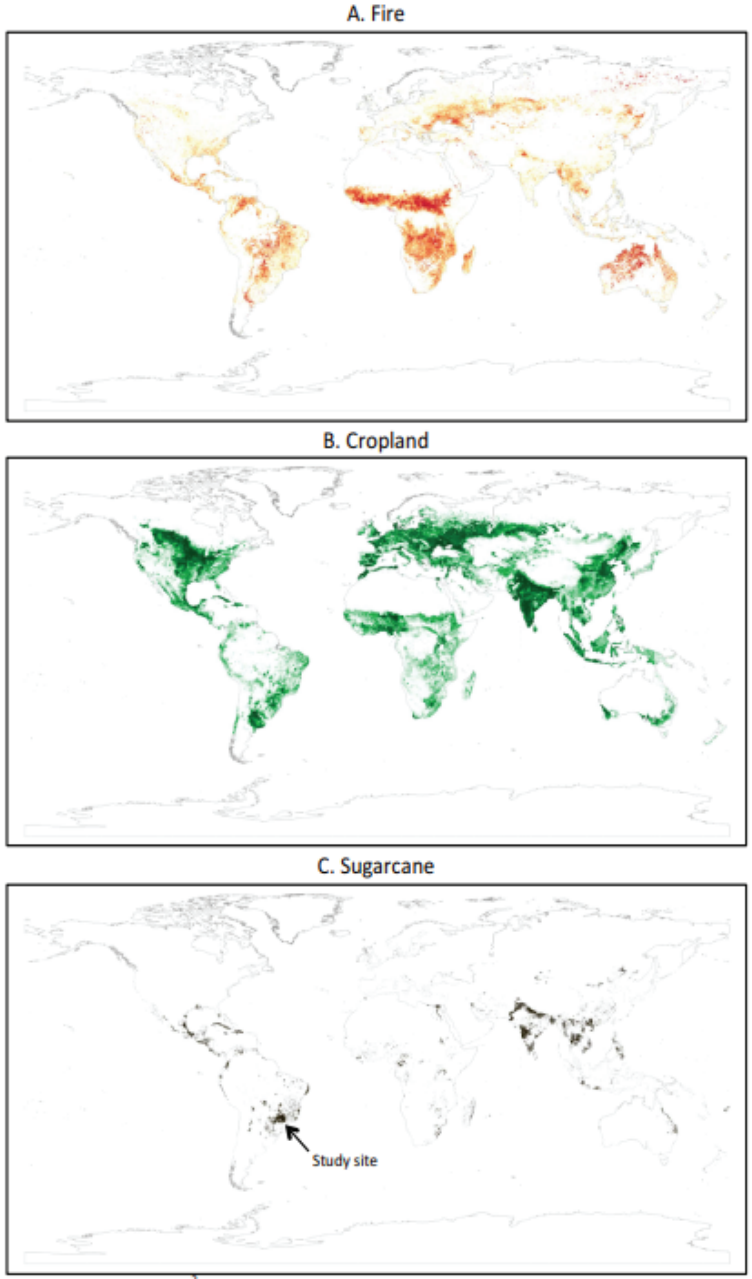
Note:This figure shows the share of sugarcane crop per municipality from 2000 to 2017. One can notice that the areas with higher concentration of this crop coincide with the areas with largest number of fire counts (C1 and C3).

Figure C9: Share of rice, sugarcane, maize and wheat planted per municipality from 2000 to 2017



Note:This figure depicts the proportion of land destined to the plantation of four crops that potentially utilise agricultural burning. One can note that sugarcane is the crop that uses this technique the most in the State of São Paulo throughout the period.

Figure C10: Distribution of Fires, cropland and sugarcane in early 2000s



Note: 5 arcminute cells, approx. 100km^2 at Equator. Darker areas indicate more fire, cropland, or sugarcane. Panel A is based on remote sensing fires data from the TERRA satellite for Nov. 2000 – Oct. 2001. Panel B is from Ramankutty et al. (2008), who estimate cropland cover circa 2000 by combining satellite remote sensing and agricultural inventory data. Panel C is from You et al. (2017), who estimate sugarcane harvest area circa 2005 using similar methods.

We base our analysis on Ramankutty, Navin, A.T. Evan, C. Monfreda, and J.A. Foley. (2008). "Farming the Planet: Geographic Distribution of Global Agricultural Lands in the Year 2000." Global Biogeochemical Cycles 22(1); and You, L., U. Wood-Sichra, S. Fritz, Z. Guo, L. See, and J. Koo. (2017). Spatial Production Allocation Model (SPAM) 2005 v3.2.

Source: Rangel and Vogl (2019). Note: This figure depicts the global distribution of fires, cropland and sugarcane. Fires relate to the period from Nov.2000 - Oct.2001. Cropland, year 2000. Sugarcane, year 2005.

Chapter 5

Conclusion

This thesis consists of three independent studies on human capital in Brazil. More precisely, I investigate the impact of educational policies, air pollution and agricultural fires on students' cognitive performance. In the first chapter, I present an outline for this thesis. In Chapter Two, I explore two educational policies as treatment variables and their impact on elementary school pupils' performance. In Chapter Three, I study the impact of air pollution in the day of the nationwide university entrance examinations on the candidates' scores. Finally, in Chapter Four, I continue to investigate the harmful effects of fetal exposure to bad air quality on fifth graders' cognitive performance, but stemming from agricultural fires from sugarcane harvesting phase.

In Chapter Two, I analyse the effect of two independent educational policies on students' performance at the national examination *Prova Brasil*, dropout and pass rates. The first policy made it optional for schools to adopt the automatic promotion from the first to the ninth grades, and the second policy increased the compulsory years of schooling from 8 to 9 years, changing the starting age from 7 to 6 years old. Regarding the first policy, I found evidence evidence suggesting that it indeed increase the pass rates as expected. However, the results indicate also a positive relationship between non-repetition in the primary school and dropout rates, which may be signalling that the policy itself

does not suffice to change the alarming dropout scenario in Brazil. With regards to the second policy which increased the compulsory years of schooling from 8 to 9 years, my findings indicate that this policy raised the Portuguese and Maths' scores in the national examinations, but also increased the dropout rates and lowered the pass rates.

These results suggest that it is not only by issuing new educational rules that the problem of low quality of education in Brazil will be sorted. For example, new staff must be trained, monitoring should apply not only the students, but also the schools, and the professionals. These remarks can be applied to other developing countries as well, with governments ensuring adequate availability of complementary tools to influence the accomplishment of the policies objectives, the efficiency and quality of primary education. Future research can cast light on the compulsory SSA policy phase, when there is data set available to assess the causality of the reform on the response variables analysed in this paper.

In Chapter Three, I investigate the contemporaneous effect of air pollution on students' scores at university entrance examination. I find evidence that bad air quality in the days of the exams hinders cognitive performance, with boys and poorer students more affected. In Chapter Four, I turn my attention to air pollution from sugarcane fires during pregnancy. My results for this research suggest that the fetal exposure to fires' smoke causes impairments in cognitive health later in life.

In Chapter Four, I explore agricultural fires from sugarcane crops in the State of São Paulo as air pollution sources. More specifically, I look at the effects of the fetal exposure to bad air quality stemming from these fires on fifth graders scores and wages later in life. Once more, I show evidence of the harmful effects of air pollution on cognitive performance, this time in the medium and long run.

These two studies shed light on the important issue of air pollution, mainly in developing countries where environmental regulations and quality of education are weaker than

in developed nations. Thus, my research topics contribute to the understanding of the effects of bad air quality in the largest developing country in the Americas. Nonetheless, as in any social sciences research, my works present a few limitations in terms of data. As for the first chapter, had we had data for the compulsory phase of the education policy that increased the years of schooling, we would have been able to identify the causal relationship between that policy and students' performance in a more accurate way. However, given that the policy is recent, there is simply no time span to assess subsequent cohorts impacted by the mandatory phase. Regarding the third and fourth chapters, a common limitation is the lack of information on the students' residence and exact place of birth, which may bring some measurement error to our analysis since we use pollution data from the municipality's population centroid as a proxy for the actual address. Yet, this is a very recurrent issue that the literature on air pollution faces. As per the fourth chapter, it would be better if we had the infants' exact date of birth. This would enable us to precisely calculate the pregnancy trimesters and, hence, reduce attenuation bias concerns. Finally, future research can extend this study by looking at the compulsory phase of the educational policies from chapter 2, and wild fires from forest burning on health and other human capital outcomes *vis-à-vis* chapter 4.

Bibliography

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19.

Abadie, A., Drukker, D., Herr, J. L., Imbens, G. W., et al. (2004). Implementing matching estimators for average treatment effects in stata. *Stata journal*, 4:290–311.

Abbey, D. E., Burchette, R. J., Knutsen, S. F., McDonnell, W. F., Lebowitz, M. D., and Enright, P. L. (1998). Long-term particulate and other air pollutants and lung function in nonsmokers. *American journal of respiratory and critical care medicine*, 158(1):289–298.

Aguiar, D. A., Rudorff, B. F. T., Silva, W. F., Adami, M., and Mello, M. P. (2011). Remote sensing images in support of environmental protocol: Monitoring the sugarcane harvest in são paulo state, brazil. *Remote Sensing*, 3(12):2682–2703.

Akagi, S., Yokelson, R. J., Wiedinmyer, C., Alvarado, M., Reid, J., Karl, T., Crounse, J., and Wennberg, P. (2011). Emission factors for open and domestic biomass burning for use in atmospheric models. *Atmospheric Chemistry and Physics*, 11(9):4039–4072.

Akerlof, G. A. and Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of economic literature*, 40(4):1167–1201.

Almond, D., Currie, J., and Duque, V. (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature*, 56(4):1360–1446.

Almond, D. and Edlund, L. (2009). Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in sweden. *The Quarterly journal of economics*, 124(4):1729–1772.

Anderson, M. L. (2015). As the wind blows: The effects of long-term exposure to air pollution on mortality. *NBER Working Paper*, (w21578).

Andreae, M. O. and Merlet, P. (2001). Emission of trace gases and aerosols from biomass burning. *Global biogeochemical cycles*, 15(4):955–966.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arbex, M. A., Cançado, J. E. D., Pereira, L. A. A., Braga, A. L. F., and Saldiva, P. H. d. N. (2004). Biomass burning and its effects on health. *J Bras Pneumol*, 30(2):158–75.

Arbex, M. A., Martins, L. C., de Oliveira, R. C., Pereira, L. A. A., Arbex, F. F., Cançado, J. E. D., Saldiva, P. H. N., and Braga, A. L. F. (2007). Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in brazil. *Journal of Epidemiology & Community Health*, 61(5):395–400.

Arceo, E., Hanna, R., and Oliva, P. (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from mexico city. *The Economic Journal*, 126(591):257–280.

Archsmith, J., Heyes, A., and Saberian, S. (2018). Air quality and error quantity: Pollution and performance in a high-skilled, quality-focused occupation. *Journal of the Association of Environmental and Resource Economists*, 5(4):827–863.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, 28(25):3083–3107.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, 122(3):1235–1264.

Banzhaf, H. S. and Walsh, R. P. (2008). Do people vote with their feet? an empirical test of tiebout. *American Economic Review*, 98(3):843–63.

Barros, R. P. (2015). Educacao e produtividade. `https://www.bndes.gov.br/wps/portal/site/home/conhecimento/seminarios/seminarios-e-apresentacoes-itens/educacao-produtividade-resumo-apresentacao`. [Online; accessed 15-November-2018].

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2):9–49.

Becker, S., Ichino, A., et al. (2002). Estimation of average treatment effects based on propensity scores. *The stata journal*, 2(4):358–377.

Bedi, A. S., Nakaguma, M. Y., Restrepo, B. J., and Rieger, M. (2020). Particle pollution and cognition: Evidence from sensitive cognitive tests in brazil. *Journal of the Association of Environmental and Resources Economists (forthcoming)*.

Bell, M. L., Peng, R. D., and Dominici, F. (2006). The exposure–response curve for ozone and risk of mortality and the adequacy of current ozone regulations. *Environmental health perspectives*, 114(4):532–536.

Berlinski, S., Galiani, S., and Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of public Economics*, 92(5-6):1416–1440.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Betthäuser, B. (2017). Fostering equality of opportunity? compulsory schooling reform and social mobility in germany. *European Sociological Review*, 33(5):633–644.

Bharadwaj, P., Eberhard, J., and Neilson, C. (2013). Health at birth, parental investments and academic outcomes. *University of California at San Diego Working Paper Series*, pages 1862–1891.

Bharadwaj, P., Eberhard, J. P., and Neilson, C. A. (2018). Health at birth, parental investments, and academic outcomes. *Journal of Labor Economics*, 36(2):349–394.

Bharadwaj, P., Gibson, M., Graff Zivin, J., and Neilson, C. (2017). Gray matters: Fetal pollution exposure and human capital formation. *Journal of the Association of Environmental and Resource Economists*, 4(2):505–542.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too young to leave the nest? the effects of school starting age. *The Review of Economics and Statistics*, 93(2):455–467.

Blau, F. D. and Kahn, L. M. (2005). Do cognitive test scores explain higher us wage inequality? *Review of Economics and statistics*, 87(1):184–193.

Blundell, R. and Costa Dias, M. (2000). Evaluation methods for non-experimental data. *Fiscal studies*, 21(4):427–468.

Bondy, M., Roth, S., and Sager, L. (2020). Crime is in the air: The contemporaneous relationship between air pollution and crime. *Journal of the Association of Environmental and Resource Economists*, 7(3):555–585.

Bruns, B., Evans, D., and Luque, J. (2011). *Achieving world-class education in Brazil: The next agenda*. The World Bank.

Buchard, V., Da Silva, A., Randles, C., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J., and Winker, D. (2016). Evaluation of the surface pm2. 5 in version 1 of the nasa merra aerosol reanalysis over the united states. *Atmospheric Environment*, 125:100–111.

Calderón-Garcidueñas, L., Engle, R., Mora-Tiscareño, A., Styner, M., Gómez-Garza, G., Zhu, H., Jewells, V., Torres-Jardón, R., Romero, L., Monroy-Acosta, M. E., et al. (2011). Exposure to severe urban air pollution influences cognitive outcomes, brain volume and systemic inflammation in clinically healthy children. *Brain and cognition*, 77(3):345–355.

Calderón-Garcidueñas, L., Mora-Tiscareño, A., Ontiveros, E., Gómez-Garza, G., Barragán-Mejía, G., Broadway, J., Chapman, S., Valencia-Salazar, G., Jewells, V., Maronpot, R. R., et al. (2008). Air pollution, cognitive deficits and brain abnormalities: a pilot study with children and dogs. *Brain and cognition*, 68(2):117–127.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1):31–72.

Cançado, J. E., Saldiva, P. H., Pereira, L. A., Lara, L. B., Artaxo, P., Martinelli, L. A., Arbex, M. A., Zanobetti, A., and Braga, A. L. (2006). The impact of sugar cane–burning emissions on the respiratory system of children and the elderly. *Environmental health perspectives*, 114(5):725–729.

Cao, J., Yang, C., Li, J., Chen, R., Chen, B., Gu, D., and Kan, H. (2011). Association between long-term exposure to outdoor air pollution and mortality in china: a cohort study. *Journal of hazardous materials*, 186(2-3):1594–1600.

Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics*, volume 3, pages 1801–1863. Elsevier.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.

Card, D. and Krueger, A. B. (1993). Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania. Technical report, National Bureau of Economic Research.

Carvalho, J. a. C. d. (2014). *Uma maior ameaça de reprovação faz os alunos estudarem mais? Uma análise do impacto da volta ao regime seriado nas escolas públicas de Ensino Fundamental: efeitos agregados e dinâmicos*. PhD thesis, University of São Paulo - Ribeirão Preto.

Casen (2006). Encuesta de caracterización socioeconómica nacional casen. `https://www.mideplan.cl/casen`. [Online; accessed 15-November-2018].

Chang, T., Graff Zivin, J., Gross, T., and Neidell, M. (2016). Particulate pollution and the productivity of pear packers. *American Economic Journal: Economic Policy*, 8(3):141–69.

Chang, T., Graff Zivin, J., Gross, T., and Neidell, M. (2019). The effect of pollution on worker productivity: evidence from call center workers in china. *American Economic Journal: Applied Economics*, 11(1):151–72.

Chay, K. Y. and Greenstone, M. (2003). The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *The quarterly journal of economics*, 118(3):1121–1167.

Chen, Huang, W., Wong, C.-M., Wang, Z., Thach, T. Q., Chen, B., Kan, H., Group, C. C., et al. (2012). Short-term exposure to sulfur dioxide and daily mortality in 17 chinese cities: the china air pollution and health effects study (capes). *Environmental research*, 118:101–106.

Chen, Li, C., Ristovski, Z., Milic, A., Gu, Y., Islam, M. S., Wang, S., Hao, J., Zhang, H., He, C., et al. (2017a). A review of biomass burning: Emissions and impacts on air quality, health and climate in china. *Science of the Total Environment*, 579:1000–1034.

Chen, S., Guo, C., and Huang, X. (2018). Air pollution, student health, and school absences: Evidence from china. *Journal of Environmental Economics and Management*, 92:465–497.

Chen, X., Zhang, X., and Zhang, X. (2017b). Smog in our brains: Gender differences in the impact of exposure to air pollution on cognitive performance.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly journal of economics*, 126(4):1593–1660.

Clougherty, J. E. (2010). A growing role for gender analysis in air pollution epidemiology. *Environmental health perspectives*, 118(2):167–176.

Cochran, W. G. and Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 417–446.

Crawford, C., Dearden, L., and Greaves, E. (2014). The drivers of month-of-birth differences in children's cognitive and non-cognitive skills. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177(4):829–860.

Crawford, C., Dearden, L., and Meghir, C. (2007). When you are born matters: The impact of date of birth on child cognitive outcomes in england.

Crawford, C., Dearden, L., and Meghir, C. (2010). When you are born matters: the impact of date of birth on educational outcomes in england. Technical report, IFS working papers.

Curi, A. Z. and Menezes-Filho, N. A. (2015). A relação entre o desempenho escolar e os salarios no brasil. Technical report, Fundação Itau Social.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47(1):87–122.

Currie, J., Graff Zivin, J., Mullins, J., and Neidell, M. (2014). What do we know about short-and long-term effects of early-life exposure to pollution? *Annu. Rev. Resour. Econ.*, 6(1):217–247.

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., and Rivkin, S. G. (2009). Does pollution increase school absences? *The Review of Economics and Statistics*, 91(4):682–694.

Currie, J. and Neidell, M. (2005). Air pollution and infant health: what can we learn from california's recent experience? *The Quarterly Journal of Economics*, 120(3):1003–1030.

Currie, J., Thomas, D., et al. (2012). Early test scores, school quality and ses: Longrun effects on wage and employment outcomes. *35th Anniversary Retrospective*, 35:185–214.

De Loecker, J. (2007). Do exports generate higher productivity? evidence from slovenia. *Journal of international economics*, 73(1):69–98.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., and Reif, J. (2019a). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, 109(12):4178–4219.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., and Reif, J. (2019b). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, 109(12):4178–4219.

Deschênes, O., Greenstone, M., and Guryan, J. (2009). Climate change and birth weight. *American Economic Review*, 99(2):211–17.

Developers, E. C. (2016). Health effects of ozone in the general population. *United States Enviromental Protection Agency*.

Dobkin, C. and Ferreira, F. (2010). Do school entry laws affect educational attainment and labor market outcomes? *Economics of education review*, 29(1):40–54.

Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris Jr, B. G., and Speizer, F. E. (1993). An association between air pollution and mortality in six us cities. *New England journal of medicine*, 329(24):1753–1759.

Drange, N., Havnes, T., and Sandsør, A. M. (2012). Kindergarten for all: Long run effects of a universal intervention. Technical report, Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit.

Ebenstein, A., Fan, M., Greenstone, M., He, G., and Zhou, M. (2017). New evidence on the impact of sustained exposure to air pollution on life expectancy from china's huai river policy. *Proceedings of the National Academy of Sciences*, 114(39):10384–10389.

Ebenstein, A., Lavy, V., and Roth, S. (2016). The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics*, 8(4):36–65.

Erickson, L. D., Gale, S. D., Anderson, J. E., Brown, B. L., and Hedges, D. W. (2020). Association between exposure to air pollution and total gray matter and total white matter volumes in adults: A cross-sectional study. *Brain Sciences*, 10(3):164.

Estadao (2019). Infográficos. `https://www.estadao.com.br/infograficos/educacao, no-enem-1-a-cada-4-alunos-de-classe-media-triunfa-pobres-sao-1-a-cada-600, 953041`. [Online; accessed 12-November-2019].

Estes Jr, M. G., Al-Hamdan, M. Z., Crosson, W., Estes, S. M., Quattrochi, D., Kent, S., and McClure, L. A. (2010). What can affect aod–pm2. 5 association? *Environmental Health Perspectives*, 118(3).

Eyles, A. and Machin, S. J. (2015). The introduction of academy schools to england's education.

FAOSTAT (2020). Fao statistics. `http://faostat.fao.org/FAOSTATFAOStatisticsDivision`. [Online; accessed 19-August-2020].

Fernandes, A. (1988). A utilização da queimada na colheita da cana-de-açúcar. *COPERSUCAR CTAG-1*, pages 1–17.

Ferrão, M. E., Beltão, K. I., and dos Santos, D. P. (2002). Políticas de não-repetência e a qualidade da educação: evidências obtidas a partir da modelagem dos dados da 4ª série do saeb-99. *Estudos em avaliação educacional*, (26):47–74.

Fertig, M. and Kluve, J. (2005). The effect of age at school entry on educational attainment in germany. *RWI*, (27). Discussion Papers.

Galizia, A. and Kinney, P. L. (1999). Long-term residence in areas of high ozone: associations with respiratory health in a nationwide sample of nonsmoking young adults [dsee comments]. *Environmental health perspectives*, 107(8):675–679.

Gan, W. Q., Koehoorn, M., Davies, H. W., Demers, P. A., Tamburic, L., and Brauer, M. (2011). Long-term exposure to traffic-related air pollution and the risk of coronary heart disease hospitalization and mortality. *Environmental health perspectives*, 119(4):501–507.

Garg, T., Jagnani, M., and Taraz, V. (2020). Temperature and human capital in india. *Journal of the Association of Environmental and Resource Economists*, 7(6):1113–1150.

Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., and Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. *Health services research*, 49(5):1701–1720.

Giglio, L., Loboda, T., Roy, D. P., Quayle, B., and Justice, C. O. (2009). An active-fire based burned area mapping algorithm for the modis sensor. *Remote sensing of environment*, 113(2):408–420.

Giglio, L., Schroeder, W., and Justice, C. O. (2016). The collection 6 modis active fire detection algorithm and fire products. *Remote Sensing of Environment*, 178:31–41.

Glewwe, P. and Kassouf, A. L. (2012). The impact of the bolsa escola/familia conditional cash transfer program on enrollment, dropout rates and grade promotion in brazil. *Journal of development Economics*, 97(2):505–517.

Gomes-Neto, J. B. and Hanushek, E. A. (1994). Causes and consequences of grade repetition: Evidence from brazil. *Economic Development and Cultural Change*, 43(1):117–148.

Goodman, A. and Sianesi, B. (2005). Early education and children's outcomes: how long do the impacts last? *Fiscal Studies*, 26(4):513–548.

Goudarzi, G., Geravandi, S., Idani, E., Hosseini, S. A., Baneshi, M. M., Yari, A. R., Vosoughi, M., Dobaradaran, S., Shirali, S., Marzooni, M. B., et al. (2016). An evaluation of hospital admission respiratory disease attributed to sulfur dioxide ambient concentration in ahvaz from 2011 through 2013. *Environmental science and pollution research*, 23(21):22001–22007.

Graff-Zivin, J., Liu, T., Song, Y., Tang, Q., and Zhang, P. (2020). The unintended impacts of agricultural fires: Human capital in china. *Journal of Development Economics*, 147:102560.

Graff-Zivin, J., Neidell, M., and Schlenker, W. (2011). Water quality violations and avoidance behavior: Evidence from bottled water consumption. *American Economic Review*, 101(3):448–53.

Guoliang, C., Zhang, X., Sunling, G., and Zheng, F. (2008). Investigation on emission factors of particulate matter and gaseous pollutants from crop residue burning. *Journal of Environmental Sciences*, 20(1):50–55.

Halla, M. and Zweimüller, M. (2014). Parental response to early human capital shocks: evidence from the chernobyl accident.

Hansen, A. C. and Selte, H. K. (2000). Air pollution and sick-leaves. *Environmental and Resource Economics*, 16(1):31–50.

Hanushek, E. A. (1996). A more complete picture of school resource policies. *Review of Educational Research*, 66(3):397–409.

Hanushek, E. A. and Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American economic review*, 90(5):1184–1208.

Heckman, J., Ichimura, H., and Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, 65(2):261–294.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, 64(4):605–654.

Herbst, M. and Strawiński, P. (2016). Early effects of an early start: Evidence from lowering the school starting age in poland. *Journal of Policy Modeling*, 38(2):256–271.

Heyes, A., Neidell, M., and Saberian, S. (2016). The effect of air pollution on investor behavior: Evidence from the s&p 500. Technical report, National Bureau of Economic Research.

Houston, J., Allendorfer, J., Nenert, R., Goodman, A. M., and Szaflarski, J. P. (2019). White matter language pathways and language performance in healthy adults across ages. *Frontiers in neuroscience*, 13.

Hsiang, S., Oliva, P., and Walker, R. (2019). The distribution of environmental damages. *Review of Environmental Economics and Policy*, 13(1):83–103.

Huber, M., Lechner, M., and Wunsch, C. (2013). The performance of estimators based on the propensity score. *Journal of Econometrics*, 175(1):1–21.

IBGE (2013). Pesquisa nacional por amostra de domicílios, pnad. *Rio de Janeiro: Instituto Brasileiro de Geografia e Estatistica, IBGE*.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1):5–86.

INEP (2017a). Microdados da prova brasil. `http://portal.inep.gov.br/web/guest/microdados`. [Online; accessed 19-January-2019].

INEP (2017b). Microdados do enem. http://portal.inep.gov.br/web/guest/microdados. [Online; accessed 19-August-2019].

IPEA (2014). Indicadores macroeconômicos. http://www.ipeadata.gov.br/Default.aspx. [Online; accessed 15-November-2018].

Isen, A., Rossin-Slater, M., and Walker, W. R. (2014). Every breath you take–every dollar you'll make: The long-term consequences of the clean air act of 1970 (no. w19858).

Itron (2019). Blogs itron. https://blogs.itron.com/computing-a-weighted-average-wind-speed-and-wind-direction-across-multi\ple-weather-stations/. [Online; accessed 19-November-2019].

Jacob, B. A. and Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of economics and statistics*, 86(1):226–244.

Jacob, B. A. and Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, 1(3):33–58.

Jayachandran, S. (2009). Air quality and early-life mortality evidence from indonesia's wildfires. *Journal of Human resources*, 44(4):916–954.

Jimerson, S., Anderson, G. E., and Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools*, 39(4):441–457.

Jimerson, S., Egeland, B., Sroufe, L. A., and Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of school psychology*, 38(6):525–549.

Justice, C., Giglio, L., Korontzi, S., Owens, J., Morisette, J., Roy, D., Descloitres, J., Alleaume, S., Petitcolin, F., and Kaufman, Y. (2002). The modis fire products. *Remote sensing of Environment*, 83(1-2):244–262.

Kahn, M. E. and Li, P. (2019). The effect of pollution and heat on high skill public sector worker productivity in china. Technical report, National Bureau of Economic Research.

Katanoda, K., Sobue, T., Satoh, H., Tajima, K., Suzuki, T., Nakatsuka, H., Takezaki, T., Nakayama, T., Nitta, H., Tanabe, K., et al. (2011). An association between long-term exposure to ambient air pollution and mortality from lung cancer and respiratory diseases in japan. *Journal of epidemiology*, pages 1102090211–1102090211.

Kaufman, Y. J., Justice, C. O., Flynn, L. P., Kendall, J. D., Prins, E. M., Giglio, L., Ward, D. E., Menzel, W. P., and Setzer, A. W. (1998). Potential global fire monitoring from eos-modis. *Journal of Geophysical Research: Atmospheres*, 103(D24):32215–32238.

Khan, R. R. and Siddiqui, M. (2014). Review on effects of particulates; sulfur dioxide and nitrogen dioxide on human health. *Int. Res. J. Environ. Sci*, 3(4):70–73.

Khandker, S., B. Koolwal, G., and Samad, H. (2009). *Handbook on impact evaluation: quantitative methods and practices*. The World Bank.

Khaniabadi, Y. O., Polosa, R., Chuturkova, R. Z., Daryanoosh, M., Goudarzi, G., Borgini, A., Tittarelli, A., Basiri, H., Armin, H., Nourmoradi, H., et al. (2017). Human health risk assessment due to ambient pm10 and so2 by an air quality modeling technique. *Process safety and environmental protection*, 111:346–354.

Kirchhoff, V., Marinho, E., Dias, P., Pereira, E., Calheiros, R., André, R., and Volpe, C. (1991). Enhancements of co and o 3 from burnings in sugar cane fields. *Journal of Atmospheric Chemistry*, 12(1):87–102.

Koppensteiner, M. F. (2010). *Essays on human capital in Brazil*. PhD thesis, Queen Mary University of London.

Koppensteiner, M. F. (2014). Automatic grade promotion and student performance: Evidence from brazil. *Journal of Development Economics*, 107:277–290.

Lai, W., Li, Y., Tian, X., and Li, S. (2018). Agricultural fires and cognitive function: evidence from crop production cycles. *Available at SSRN 3039935*.

Lamsal, K., Jones, P. C., and Thomas, B. W. (2017). Sugarcane harvest logistics in brazil. *Transportation Science*, 51(2):771–789.

Lavy, V., Paserman, M. D., and Schlosser, A. (2011). 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):208–237.

Lechner, M. (2002). Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(1):59–82.

Leuven, E. and Sianesi, B. (2003). Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components S*, 432001.

Lin, C.-M., Li, C.-Y., Yang, G.-Y., and Mao, I.-F. (2004). Association between maternal exposure to elevated ambient sulfur dioxide during pregnancy and term low birth weight. *Environmental research*, 96(1):41–50.

Ling, S. H. and van Eeden, S. F. (2009). Particulate matter air pollution exposure: role in the development and exacerbation of chronic obstructive pulmonary disease. *International journal of chronic obstructive pulmonary disease*, 4:233.

Lipfert, F. W., Wyzga, R. E., Baty, J. D., and Miller, J. P. (2009). Air pollution and survival within the washington university–epri veterans cohort: risks based on modeled estimates of ambient levels of hazardous and criteria air pollutants. *Journal of the Air & Waste Management Association*, 59(4):473–489.

Liu, H. and Salvo, A. (2018). Severe air pollution and child absences when schools and parents respond. *Journal of Environmental Economics and Management*, 92:300–330.

Liu, Y., Paciorek, C. J., and Koutrakis, P. (2009). Estimating regional spatial and temporal variability of pm2. 5 concentrations using satellite data, meteorology, and land use information. *Environmental health perspectives*, 117(6):886–892.

Machin, S., Marie, O., and Vujić, S. (2011). The crime reducing effect of education. *The Economic Journal*, 121(552):463–484.

Machin, S. and Wilson, J. (2005). Public and private schooling initiatives in england. In *Conference on Mobilizing the Private Sector for Public Education, October*, pages 5–6.

Makles, A. (2012). Stata tip 110: How to get the optimal k-means cluster solution. *The Stata Journal*, 12(2):347–351.

Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics*, 94(2):596–606.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, 107(2):407–437.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Marques da Costa, R. M. V. (2018). *Education, Work and Crime: Evidence from Educational Reforms*. PhD thesis, University College London.

McConnell, M., Dohlman, E., and Haley, S. L. (2010). World sugar price volatility intensified by market and policy factors. Technical report.

McEwan, P. J. and Shapiro, J. S. (2008). The benefits of delayed primary school enrollment discontinuity estimates using exact birth dates. *Journal of human Resources*, 43(1):1–29.

Moretti, E. and Neidell, M. (2011). Pollution, health, and avoidance behavior evidence from the ports of los angeles. *Journal of human Resources*, 46(1):154–175.

Murnane, R. J. and Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Murnane, R. J., Willett, J. B., and Levy, F. (1995). The growing importance of cognitive skills in wage determination. Technical report, National Bureau of Economic Research.

Murphy, K. M. and Peltzman, S. (2004). School performance and the youth labor market. *Journal of labor Economics*, 22(2):299–327.

Neidell, M. (2004). Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma. *Journal of health economics*, 23(6):1209–1236.

Neidell, M. (2006). Public information and avoidance behavior: Do people respond to smog alerts? *Center for Integrating Statistical and Environmental Science Technical Report*, 24.

Neidell, M. (2009). Information, avoidance behavior, and health the effect of ozone on asthma hospitalizations. *Journal of Human resources*, 44(2):450–478.

Nilsson, J. P. (2009). The long-term effects of early childhood lead exposure: Evidence from the phase-out of leaded gasoline. *Institute for Labour Market Policy Evaluation (IFAU) Work. Pap.*

Okurut, J. M. (2018). Automatic promotion and student dropout: Evidence from uganda, using propensity score in difference in differences model. *Journal of Education and Learning*, 7(2):191.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1):152–175.

Paciorek, C. J. and Liu, Y. (2009). Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter. *Environmental health perspectives*, 117(6):904–909.

Peña, P. A. (2017). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review*, 56:152–176.

Popescu, T., Sader, E., Schaer, M., Thomas, A., Terhune, D. B., Dowker, A., Mars, R. B., and Kadosh, R. C. (2019). The brain-structural correlates of mathematical expertise. *cortex*, 114:140–150.

Provençal, S., Buchard, V., da Silva, A. M., Leduc, R., and Barrette, N. (2017). Evaluation of pm surface concentrations simulated by version 1 of nasa's merra aerosol reanalysis over europe. *Atmospheric pollution research*, 8(2):374–382.

Rangel, M. A. and Vogl, T. S. (2019). Agricultural fires and health at birth. *Review of Economics and Statistics*, 101(4):616–630.

Ransom, M. and Pope, C. (2013). Air pollution and school absenteeism: Results from a natural experiment. *Manuscript, Department of Economics, Brigham Young University.*

Reyes, J. W. (2011). Childhood lead and academic performance in massachusetts. *Childhood*, 11(3).

Ribeiro, H. (2008). Sugar cane burning in brazil: respiratory health effects. *Revista de Saúde Pública*, 42:370–376.

Rivkin, S. G. (1995). Black/white differences in schooling and employment. *Journal of Human resources*, pages 826–852.

Roderick, M. (1994). Grade retention and school dropout: Investigating the association. *American Educational Research Journal*, 31(4):729–759.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):S71–S102.

Rosa, L., Martins, M., and Carnoy, M. (2019). Achievement gains from reconfiguring early schooling: The case of brazil's primary education reform. *Economics of Education Review*, 68:1–12.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Rubin, D. B. (1980). Bias reduction using mahalanobis-metric matching. *Biometrics*, pages 293–298.

SAEB (2006). Escala saeb. ministerio da educacao. `https://www.academia.qedu.org.br/prova-brasil/454-2/`. [Online; accessed 15-November-2018].

Sager, L. (2019). Estimating the effect of air pollution on road safety using atmospheric temperature inversions. *Journal of Environmental Economics and Management*, 98:102250.

Sanders, N. J. (2012). What doesn't kill you makes you weaker prenatal pollution exposure and educational outcomes. *Journal of Human Resources*, 47(3):826–850.

Schlenker, W. and Walker, W. R. (2015). Airports, air pollution, and contemporaneous health. *The Review of Economic Studies*, 83(2).

Seaton, A., Godden, D., MacNee, W., and Donaldson, K. (1995). Particulate air pollution and acute health effects. *The lancet*, 345(8943):176–178.

SEB/MEC (2006). Ensino fundamental de nove anos: perguntas mais frequentes e respostas da secretaria de educação básica (seb/mec) kernel description. `http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm`. [Online; accessed 13-November-2018].

Setzer, A. and Sismanoglu, R. (2007). Risco de fogo: Metodologia do calculo. *Instituto Nacional de Pesquisas Espaciais*.

Sianesi, B. (2001). Implementing propensity score matching estimators with stata. In *UK Stata Users Group, VII Meeting*.

SIMA (2017). The green protocol. secretariat for infrastructure environment state of são paulo. https://www.infraestruturameioambiente.sp.gov.br/en/greener-ethanol/. [Online; accessed 10-November-2019].

Smith, J. (1996). Three essays on the economics of evaluating social programs: Dissertation summary.

Solow, R. M. (1999). Neoclassical growth theory. *Handbook of macroeconomics*, 1:637–667.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56(1):79–107.

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., Danabasoglu, G., Suzuki, T., Bamber, J. L., Bentsen, M., et al. (2018). Jra-55 based surface dataset for driving ocean–sea-ice models (jra55-do). *Ocean Modelling*, 130:79–139.

UN (2015). Millenium development goals. http://www.un.org/millenniumgoals/. [Online; accessed 18-November-2018].

UNESCO, E. f. A. P. (2007). *Education for All Global Monitoring Report 2008: Education for All by 2015. Will We Make It?* Oxford University Press.

Unicadata (2020). Observatório da cana. http://www.unicadata.com.br. [Accessed in August, 2020].

Uriarte, M., Yackulic, C. B., Cooper, T., Flynn, D., Cortes, M., Crk, T., Cullman, G., McGinty, M., and Sircely, J. (2009). Expansion of sugarcane production in são paulo, brazil: implications for fire occurrence and respiratory health. *Agriculture, ecosystems & environment*, 132(1-2):48–56.

Vanino, E. (2016). *Essays on firms innovation, internationalization and trade policy*. PhD thesis, University of Birmingham.

Webmet (2019). The meteorological resource centre. http://www.webmet.com/met_monitoring/622.html. [Online; accessed 11-November-2019].

Weinhold, B. (2008). Ozone nation: Epa standard panned by the people. https://ehp.niehs.nih.gov/doi/full/10.1289/ehp.116-a302. [Online; accessed 19-August-2019].

Wikipedia (2007). Municipalities of brazil. https://en.wikipedia.org/wiki/Municipalities_of_Brazil. [Online; accessed 15-November-2018].

Wikipedia (2019). Ozone. https://en.wikipedia.org/wiki/Ozone#Ozone_air_pollution. [Online; accessed 16-November-2019].

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wu, J., Kong, S., Wu, F., Cheng, Y., Zheng, S., Yan, Q., Zheng, H., Yang, G., Zheng, M., Liu, D., et al. (2018). Estimating the open biomass burning emissions in central and eastern china from 2003 to 2015 based on satellite observation. *Atmospheric Chemistry & Physics*, 18(16).

Zapata, J., Pont, B., Figueroa, D. T., Peterka, J., and Fraccola, S. (2015). Education policy outlook: Brazil. *OECD Publishing*.

Zhang, L., Liu, Y., and Hao, L. (2016). Contributions of open crop straw burning emissions to pm2. 5 concentrations in china. *Environmental Research Letters*, 11(1):014014.

Zhang, S., Zhong, R., and Zhang, J. (2017). School starting age and academic achievement: Evidence from china's junior high schools. *China Economic Review*, 44:343–354.

Zweig, J. S., Ham, J. C., and Avol, E. L. (2009). Air pollution and academic performance: Evidence from california schools. *Department of Economics, University of Maryland*.