Essays in Applied Microeconomics

María Lombardi

TESI DOCTORAL UPF / ANY 2017

DIRECTORS DE LA TESI
Libertad González i Stephan Litschig

Departament d’Economia i Empresa

Acknowledgements

I am extremely grateful to Libertad and Stephan for all their support throughout these years. Your advice and constant encouragement has meant a great deal. I am also indebted to Gianmarco for trusting and guiding me from early on, and to Alessandro and everyone in the UPF LPD Group. I would also like to thank Marta and Laura for their patience and endless help.

I almost entirely owe the girls from Aragó 347 and my group of friends from Pompeu for making these years in Barcelona so memorable. También quiero agradecer a Laura y Kike, que más que mis tíos fueron como mis papás (o hermanos) durante estos años. A las chicas de Buenos Aires, gracias por siempre hacerme sentir como si nunca me hubiese ido. También quiero agradecer a Gigio, que supuestamente me convenció de venir a Barcelona. Finalmente, a mi mamá y mi papá, claramente no hubiese podido hacer ni un poquito de todo esto sin su apoyo y cariño. ¡Gracias!

Abstract

This thesis is composed of three independent articles. The first chapter studies a nationwide teacher pay-for-performance program in Peru, and shows that it had a precisely estimated null impact on student learning. I provide suggestive evidence that some of the program's characteristics might have hindered teachers' ability to improve the incentivized outcome or infer their probability of winning the bonus, explaining this zero effect. In the second chapter, I examine the impact of compulsory voting laws in Austria, and show that while these laws led to significant increases in turnout, they had no impact on government spending patterns or electoral outcomes. Individual-level data suggests these results occur because individuals swayed to vote due to compulsory voting are more likely to be non-partisan and uninformed, and have less interest in politics. In the final chapter, I study how income inequality affects growth in Brazil, and find that places with higher initial inequality exhibit higher subsequent growth. I show that this effect is entirely driven by inequality originating in the left tail of the income distribution, and provide evidence on channels consistent with credit constraints and setup costs for human and physical capital investments, as well as an increasing and concave individual propensity to save.

Resumen

Esta tesis se compone de tres artículos independientes. El primer capítulo estudia un programa nacional de remuneración basado en el desempeño en el Perú, y muestra que tuvo un efecto nulo y precisamente estimado sobre el aprendizaje estudiantil. Proporciono evidencia sugiriendo que algunas de las características del programa podrían haber limitado la capacidad de los docentes de mejorar en términos del indicador incentivado o inferir su probabilidad de ganar el bono, explicando así este efecto nulo. En el segundo capítulo, examino el impacto de las leyes de voto obligatorio en Austria, y muestro que aunque estas incrementaron considerablemente la participación electoral, no tuvieron impacto alguno sobre el gasto público o el resultado de las elecciones. Un análisis de información a nivel individual sugiere que estos resultados se dan porque los individuos que votan a

raíz de su obligatoriedad son menos propensos a tener una afiliación partidaria, y están menos informados e interesados en la política. En el último capítulo estudio cómo la desigualdad de ingresos afecta al crecimiento económico en Brasil, y encuentro que aquellos lugares con mayor desigualdad inicial crecen más. Muestro que este efecto está impulsado íntegramente por la desigualdad originada en la parte inferior de la distribución de ingresos, y presento evidencia sobre mecanismos consistente con la existencia de restricciones crediticias e indivisibilidad de la inversión en capital físico y humano, así como una propensión individual al ahorro creciente y cóncava.

Preface

This doctoral thesis combines three self-contained essays, cutting across the fields of education, political economy and development. While independent in nature, a common feature of these articles is their empirical approach and the fact that they are rooted in the relevant theoretical literature.

In most education systems, compensation is based on lifetime job tenure, and teachers' salaries tend to be flat and linked to seniority, providing weak incentives for excellence in teaching. The first chapter, coauthored with Cristina Bellés Obrero, examines whether tying teachers' pay to the performance of students can increase student learning. We study this in the context of *Bono Escuela* (BE), a nationwide teacher pay-for-performance program implemented in 2015 in all public secondary schools in Peru. Taking advantage of a feature of the program by which a schools' probability of obtaining the BE bonus hinges on the achievement of 8^{th} grade students in 2015, we perform a difference-in-differences estimation, comparing the change in performance of 8^{th} grade students with that of 9^{th} graders attending the same school, before and after the implementation of BE. Despite the large monetary incentives provided by BE, we find that the program had a precisely estimated zero effect on student achievement. We provide suggestive evidence that the null effect we uncover can be explained by some of the program's characteristics, which might have made teachers unable to react in such a short time frame, unsure about their potential ranking within the group of schools they were competing against, or uncertain about how to influence their students' performance in the standardized test tied to the bonus.

The second chapter of this thesis studies the impact of compulsory voting (CV) laws on electoral participation, election results and government spending. Despite the centrality of elections to democracy, many people fail to vote, potentially leading disadvantaged groups, who are less likely to vote, to be under-served by their government. While CV might help address this issue, there is no empirical evidence on whether it affects government policy. Together with Mitchell Hoffman and Gianmarco León, we study a unique quasi-experiment in Austria, where CV laws are changed across Austria's nine states at different times. Using state-level voting records on state and national elections from 1949-2010, we find that CV increases turnout in parliamentary, state and presidential elections from roughly 80% to 90%. However, we find that CV has no impact on state-level spending, either in levels or in composition, or on electoral outcomes. We provide evidence on mechanisms by analyzing the interaction effects of CV laws with voter characteristics. Using repeated cross-sections of individual level data, we find suggestive evidence of larger CV impacts among those who have low interest in politics, have no party affiliation, and are relatively uninformed. While suggestive, these results are consistent with a story where voters who vote or abstain due to the introduction or repeal of CV do not have strong policy or partisan preferences, thereby having little or no effect on electoral outcomes. Furthermore, if such voters are unresponsive to policy in deciding which party to support, parties may have little incentive to shape policies to suit those voters' preferences.

The third chapter, coauthored with Stephan Litschig, studies whether inequality arising in the lower as opposed to the upper tail of the income distribution has different effects on subsequent income per capita growth. Greater inequality as measured by commonly used metrics can result from higher dispersion in different parts of the income distribution, and theory suggests that its growth impacts might differ depending on which tail it originates from. However, empirical work has effectively treated the variation in overall inequality the same. Using sub-national data from Brazil over the 1970-2000 period, we first establish that controlling for initial income per capita and a host of standard confounders, places with higher initial income inequality as measured by the Gini coefficient exhibit higher subsequent income per capita growth. We then propose a simple approach to distinguish between growth effects of inequality originating from the bottom versus the top of the initial income distribution. By including quintile income shares instead of the Gini coefficient in an otherwise standard cross-sectional growth regression, we allow for hypothetical income redistributions from the two tails towards the (omitted) middle quintile, while holding other income shares and mean income constant. We find that the positive association between overall initial inequality and subsequent growth is entirely driven by inequality in the lower tail of the income distribution. Consistent with explanations based on credit market and savings channels, we show that places with higher inequality in the left tail also accumulation physical and human capital at a faster pace.

Contents

List of Figures

List of Tables

Chapter 1

Teacher Performance Pay and Student Learning: Evidence from a Nationwide Program in Peru

Joint with Cristina Bellés Obrero (Universitat Pompeu Fabra)

1.1 Introduction

Teacher quality is one of the key factors determining student achievement. Individuals exposed to better teachers not only perform better in school (Hanushek and Rivkin, 2010; Rockoff, 2004; Araujo et al., 2016), but are also more likely to attend college and earn higher salaries (Chetty et al., 2014). However, the payment schemes in most educational systems do not provide adequate incentives for excellence in teaching. With relatively flat salary progression, promotion policies rigidly linked to seniority, and lifetime job tenure, these types of compensation policies might discourage high skilled individuals from taking on the teaching profession, and create weak incentives for existing teachers to exert high levels of effort (Bruns et al., 2011). In an attempt to increase teacher motivation, accountability, effort, and ultimately student learning, academics and policymakers have proposed tying teachers' pay to their students' performance. Pay-for-performance programs in education have been implemented in high income countries like United States, England, the Netherlands and Israel,[1] as well as in developing countries such as India, Pakistan, Kenya, China, Chile, Brazil, Mexico, and more recently in Peru. However, the evidence on the effectiveness of teacher incentives is scant and inconclusive; only part of these programs have been rigorously evaluated, and those that have been studied often differ in their conclusions.

This paper studies the impact of *Bono Escuela* (BE) on student achievement. BE is a nationwide teacher pay-for-performance program implemented in 2015 in public secondary schools in Peru. The program takes the form of a rank-order tournament in which all Peruvian public secondary schools compete within a group of comparable schools on the basis of their annual performance. Every teacher and the principal in schools ranked in the top 20% within their BE group obtain a fixed payment amounting to over a month's salary. The incentives provided by BE are collective (at the school-level), as all teachers are rewarded if their school wins, although the main performance measure used to rank schools is their average score in the 2015 nationwide math and language standardized tests, taken only by 8^{th} graders. This feature of the program, which we exploit in our identification, implies that a school's probability of obtaining the bonus hinges on the achievement of 8^{th} grade students in 2015. Our estimation relies on a novel administrative database collected by the Peruvian Ministry of Education, which covers the universe of students in 2013-2015 and contains annual information on the grades that students receive from their teachers in each subject (their "internal grades").[2] Importantly, teachers' grading tactics should not be influenced by the incentive, since internal grades have no direct impact on a school's BE score. We provide ample evidence that internal grades are correlated with standardized measures of learning, and show that the same teacher typically grades students from parallel classes differently, suggesting that grading on a curve is not the norm in Peruvian secondary schools. The availability of achievement measures for students in all grades allows us to compare changes in the internal grades of 8^{th} graders to those of 9^{th} grade students attending the same school, before and after

[1]An exhaustive list of OECD countries with teacher pay-for-performance programs is provided in OECD Education at a Glance 2011, available at https://www.oecd.org/education/skills-beyond-school/48631582.pdf

[2]We borrow this terminology from Calsamiglia and Loviglio (2016).

the incentive was introduced, providing difference-in-differences estimates of the effect of BE on student achievement.

While providing teachers with extrinsic monetary incentives could encourage them to exert more effort, positively impacting student performance, these incentives might yield no improvement if the incentives are not large enough, not understood, or if teachers do not know how to increase student achievement, for example. Teacher incentives might also be ineffective, or even detrimental to student learning, if they lead teachers to engage in undesirable practices such as targeting topics likely to be tested, coaching students on test-taking strategies, or cheating.[3] Student learning could also decrease if the program crowds out teachers' intrinsic motivation.[4] Since a school's probability of obtaining the BE bonus depends on the effort of many of its teachers, the program has the potential of inducing teacher free-riding (Holmstrom, 1982), thus lowering its impact on student learning. However, rewarding the entire school might have the benefit of promoting higher cooperation and monitoring among teachers (Kandel and Lazear, 1992; Kandori, 1992).

We find that the program had no impact on students' math and language internal grades. Our coefficients are precisely estimated, allowing us to reject effects larger than 0.017 standard deviations (SD) in both math and language, well below the treatment effects found in the existing literature (around 0.15-0.25 SD). Furthermore, when separately examining the impact of the program in each of the 395 groups in which schools compete, we find zero average effects in the majority of cases, providing additional evidence of the null effect of BE on student achievement. Using data on the overlap of teachers in 8^{th} and 9^{th} grade, we assess whether BE generated improvements in student achievement in our comparison group, and discard the existence of any spillovers which could bias our estimates.

[3]This type of behavior is consistent with models of multi-tasking (Holmstrom and Milgrom, 1991; Baker, 1992, 2002), and has been reported in several studies on teacher incentives such as Jacob and Levitt, 2003, Figlio and Winicki, 2005, Figlio, 2006, Glewwe et al., 2010, and Behrman et al., 2015, to name a few. Although these actions could still improve the performance of students whose teachers were devoting little time to effective teaching, they might not affect student learning, or even harm it if they crowd-out effective instruction time (Koretz, 2002; Neal, 2011).

[4]Previous research shows that monetary incentives produce not only a price effect, making the incentivized behavior more attractive, but also a psychological effect that can crowd out the former. There is evidence of this behavior, albeit in a different context, in the studies of Gneezy and Rustichini (2000a), Gneezy and Rustichini (2000b) and Gneezy et al. (2011).

Our null average treatment effects could be masking the fact that teachers from some schools might be more incentivized than other, due to the tournament nature of BE. In particular, the incentive could be impacting schools closer to the margin, and leaving sure-winners and sure-losers unaffected (Contreras and Rau, 2012). We explore whether there are differential effects across these and other dimensions of the incentive, and do not find evidence of heterogeneous effects.

Why was student learning unaffected by the BE program? In order to answer this question, we carried out an online survey on a sample of public secondary school teachers regarding the 2015 BE. We provide suggestive evidence that the null effect is not a result of the size or collective nature of the incentive, or driven by teachers being uninformed about the BE, or only focusing on increasing standardized test scores -the incentivized outcome- without influencing their students' learning in a meaningful way. We put forth a few reasons why the program may have had a null effect. Firstly, certain features of the standardized test linked to the bonus might have hampered teachers' ability to boost student performance in terms of this measure, potentially discouraging teachers from exerting higher effort. Given that students were tested for the first time in 2015, teachers might not have known what pedagogical practices result in higher test scores. The fact that students have no stakes in these evaluations might also have played a role in weakening the mapping between teachers' effort and their chances of winning the bonus. Secondly, the incentive might have been diluted if schools were uncertain about their potential ranking within the group of schools they were competing against. Given that they had no prior experience with the standardized test tied to BE, this is not unlikely. Finally, we argue that teachers might not have had enough time to react to the incentive. The future analysis of the 2016 wave of the BE, for which some of these issues will be alleviated, will also allow us to better pin-down these channels.

Our paper relates to the literature on teacher pay-for-performance, particularly in the context of other developing countries. Although a few studies find positive and significant effects on student learning, the literature reveals mixed results. Using a randomized controlled trial in rural schools in the Indian state of Andhra Pradesh, Muralidharan and Sundararaman (2011) study the effect of providing individual and collective monetary incentives to teachers based on students' test score improvements. The incentive had a significant and sizable effect on students' standardized test scores and a positive impact on other subjects not targeted by the incentive. The experimental study of Glewwe et al. (2010) evaluates a collective teacher incentive program in Kenya, and finds that although the program yielded a positive effect on students' test scores in the exams tied to the incentive, it had no impact on non-incentivized exams covering similar topics.[5] Behrman et al. (2015) implement a randomized controlled trial in a sample of Mexican high schools, providing monetary incentives to teachers and/or students based on the latter's performance in math tests with very low stakes. The authors find that, while providing monetary incentives to teachers had no impact on students' math test scores, there was a significant increase in student performance when students themselves were incentivized. The effects were larger when both students and teachers were given incentives.[6] Barrera-Osorio and Raju (2017) evaluate a performance-pay program in Pakistan giving bonuses to primary school principals and/or teachers according to their school's increase in enrollment rates, and the participation rate and scores of their students in a standardized test. The authors only find an increase in the exam participation rates, and argue that this is consistent with the incentives introduced by the program. The closest paper to ours is that of Contreras and Rau (2012), who examine the impact of a scaled-up program in Chile. Using matching and difference-in-differences techniques, they find that public school students performed significantly better in math and language as compared to students in private schools, which were not eligible for the bonus.

In the context of a high income country, the quasi-experimental studies of

[5]Although direct observation of teachers in Muralidharan and Sundararaman (2011) shows no impact of the teacher incentive on classroom processes, or student and teacher attendance, teachers in treatment schools were more likely to report having assigned extra homework, classwork and practice tests, conducting extra classes, and paying special attention to weaker students. External observers in Glewwe et al. (2010) also found no changes in teacher attendance, homework assignment or pedagogy. However, the principals of treated schools were more likely to report that their teachers offered extra prep classes, suggesting that teachers' efforts might have narrowly targeted the incentivized outcome.

[6]Consistent with the authors' findings, incentivized students reported exerting higher effort in preparing for the exam. Self-reported behavior of teachers is not as compatible with the results, since teachers in all three treatment arms were more likely to report having prepared their students for the test, both inside and outside of class.

Lavy (2002) and Lavy (2009) in Israeli high schools examine a collective and an individual teacher incentive program, respectively, and find positive and significant impacts in different measures of student performance tied to the incentives. In a follow-up paper, Lavy (2015) finds that ten years after the pay-for-performance program examined in Lavy (2009), students from treated schools exhibited significantly higher level of schooling attainment and higher wages. Fryer (2013), and Goodman and Turner (2013) independently analyze a randomized controlled experiment in over 200 New York City public schools where schools meeting their performance target could earn a lump sum payment, which they could distribute at their own discretion. Both studies find no evidence of increased student attainment or changes in students' or teachers' behavior. Finally, Springer et al. (2010) conducted a three-year study in the Metropolitan Nashville School System in which math teachers were economically incentivized for large gains on standardized tests, and find a positive effect only among teachers instructing the same set of students in multiple subjects.

Performance pay schemes have traditionally been examined in the context of organizations.[7] There have been several studies examining the causal effect of linking managerial pay to overall firm performance (Groves et al., 1994, Chevalier and Ellison, 1997, and Oyer, 1998, among others) or to the productivity of bottom-tier workers (Bandiera et al., 2007). Other papers have focused instead on the impact of different payment schemes on worker productivity (e.g., Bandiera et al., 2005 and Bandiera et al., 2013). The results of our study can be informative for this other stream of the literature as well.

One contribution of our paper is that we examine whether teacher pay-for-performance can work in the context of a scaled-up, national intervention. Except for Contreras and Rau (2012), all the other studies in this literature tackle this question using a randomized controlled trial, in which the scale is necessarily smaller, and responsibility for the implementation is usually handed over to an NGO (instead of the government). For instance, Barrera-Osorio and Raju (2017) evaluate a pay-for-performance program implemented in 450 schools,[8]

[7] See Prendergast (1999) for a general review of the early empirical evidence on the provision of incentives in firms.

[8] Although in comparison to BE this program was implemented in a reduced sample of schools, it shares the feature that it was designed and managed by the government.

Muralidharan and Sundararaman (2011) in 200 schools, Glewwe et al. (2010) in 50 schools, and Behrman et al. (2015) in 40 schools. In contrast, the BE program was implemented solely by the Peruvian Ministry of Education, and reached more than 8,000 schools across Peru, providing incentives to roughly 81,000 teachers, responsible for instructing 70% of Peruvian students in the 8^{th} grade. While these experimental studies make important contributions towards understanding whether teacher pay-for-performance can increase student achievement, they face external validity issues, as in any randomized controlled trial (Deaton and Cartwright, 2016), making their findings not necessarily generalizable to a large-scale program. This notion is put forward in Banerjee et al. (2016), where a successful educational intervention led by a NGO did not yield the same initial impact when it was scaled-up and implemented within the existing educational system. Budgetary constraints (Kerwin and Thornton, 2015) and opposition from teacher unions (Bruns and Luque, 2015; Mizala and Schneider, 2014) make several aspects of these types of interventions unfeasible in a nationwide program. For example, students in Muralidharan and Sundararaman (2011) were evaluated at baseline, and their teachers received feedback on their performance in each question. Testing students so often and providing such detailed feedback to their teachers might be too costly to implement on a national scale. It is thus crucial for policymakers to better understand the role played by the features of teacher pay-for-performance programs. While we cannot fully tease out which characteristics of the BE contributed to its null impact, we provide some suggestive evidence, hopefully shedding more light on this discussion.

Another novelty of our study is its use of a measure of student achievement that captures the skills of students which are targeted by the program without being directly incentivized. Since the BE bonus is linked to standardized test scores and not to internal grades, teachers' stakes in our outcome variable are not modified by the incentive.[9] An identification strategy relying on standardized test scores (or other incentivized indicators) as an outcome cannot fully disentangle whether improvements in students' performance are the consequence of higher learning or

[9] Students' internal grades in Peru are completely independent of standardized test scores. For one, standardized tests are graded after the end of the school year, and students' individual scores are never reported.

the results of short-term strategies fostering high test scores (Neal, 2011). The importance of this issue is highlighted by the results from Glewwe et al. (2010), who find that while students performed better in the tests used to award the bonus, there was no effect in their performance in an alternative exam not linked to the incentive. With the exception of the latter, all the other papers in this literature assess student achievement using measures of learning which are directly targeted by the incentive.[10] While it would also be interesting to analyze the impact of BE on standardized test scores, it is not possible because 2015 was the first year in which these tests were applied in secondary schools, and there is no appropriate comparison group. Using internal grades as our outcome has the advantage of capturing students' performance without directly influencing teachers' probability of obtaining the bonus. While this measure might have some shortcomings, for instance if teachers assign grades on a relative basis, we report the results from multiple tests alleviating this concern.

The paper is organized as follows. Section 1.2 describes the educational system in Peru and the Peruvian teacher pay-for-performance program. Section 1.3 discusses our strategy for estimating the effect of teacher incentives on student performance, and Section 1.4 describes the data. Section 1.5 presents our main results, and Section 1.6 provides evidence on the validity of our identification assumption. Section 1.7 discusses the potential reasons for our findings, and Section 1.8 concludes.

[10] Behrman et al. (2015) and Contreras and Rau (2012) measure achievement using students' scores in the standardized evaluations tied to the bonus. The studies of Fryer (2013) and Goodman and Turner (2013) use several measures of student performance linked to the incentive (scores in state tests, graduation rates, credits earned, etc.), as do Lavy (2002) and Lavy (2009) (average score and pass rates in matriculation exams, and school dropout rates). In an attempt to overcome this issue, Muralidharan and Sundararaman (2011) designed the standardized test to include mechanical and conceptual questions; while performance in the former can be easily affected by a teacher coaching his students for the test, conceptual questions are harder to influence using these types of tactics.

1.2 Secondary Schooling in Peru and the BE Program

1.2.1 Secondary Schooling in Peru

Compulsory schooling is 12 years in Peru, and is composed of initial, primary and secondary schooling, lasting one, six, and five years. Students in public secondary schools have seven hours of instruction a day, although the Ministry of Education has been gradually implementing nine-hour school days, currently reaching 18% of all public secondary schools. While a significant portion of the student body attends private secondary schools, public institutions dominate by far. In 2014, for instance, 63% of high schools were publicly run, instructing 76% of all secondary students. Over the last decade there have been significant improvements in secondary school coverage, with enrollment rising from 71% of individuals in secondary school age (12-16) in 2005 to 83% in 2014. Despite these improvements, enrollment is still far from universal. Moreover, a very high portion of students attending high school do not possess the minimum required levels of knowledge. In the 2012 round of OECD's Programme for International Student Assessment (PISA) evaluating 15-year-old students, Peru was the lowest scoring country out of 65 in all three tested subjects. In particular, 75%, 60% and 69% of Peruvian students had low achievement in math, reading and science, respectively.

Public school teachers in Peru can be either civil servants or contract teachers. Salaries for the former are divided into eight pay scales, with a monthly salary of 1451 soles ($\approx$ 439 dollars) in the lowest scale in 2015, and a salary of 3773 soles ($\approx$ 1142 dollars) in the highest.[11] Contract teachers, on the other hand, received a fixed monthly payment of 1244 soles ($\approx$ 370 dollars).[12] There were approximately 120,000 public secondary school teachers in 2014, one third of which were contract teachers. The average secondary school teacher in public schools received a monthly salary of only 1469 soles, roughly 444 dollars.[13] The working week

[11] Throughout this study we use a conversion rate of 3.31 soles per dollar.

[12] Further details on teachers' salaries and pay scales are provided by the Ministry of Education in http://www.minedu.gob.pe/reforma-magisterial/remuneraciones-beneficios.php, last accessed August 16, 2016.

[13] We calculated the average monthly salary of public secondary school teachers in Peru using

for public secondary school teachers in 2015 consisted of 26 hours, 24 of which were to be spent teaching.[14] However, as reported in a nationally representative teacher survey at the end of 2014, teachers spent an average of 12 hours a week performing other activities outside their official working hours, such as preparing class materials or attending parent-teacher conferences. Furthermore, 15% of secondary school teachers taught in more than one school, and 28% complemented their salary with another type of job.

According to this same survey, 52% of public secondary school teachers had a university degree, 45% obtained their teaching degree in a tertiary institution, and the remaining 3% had another type of degree, or no degree at all. As compared to Peruvian workers with similar qualifications, and teachers in comparable countries, Peruvian teachers are poorly paid. A study on Latin American teachers' salaries in 2010 shows that adjusting for the number of hours worked, Peruvian teachers made 10% less than other Peruvian professional workers with similar education (Bruns and Luque, 2015). In comparison to individuals with similar qualifications, teachers in Peru were paid relatively worse than in Mexico, Honduras, El Salvador, Costa Rica, Uruguay and Chile, but relatively better than than in Panama, Brazil and Nicaragua.[15] While Peruvian teachers are poorly paid, absenteeism is quite low. Around the time in which BE was implemented, teacher absenteeism was below 7% in public schools around the country.[16]

the Ministry of Education's pay scales and the type of contract and category reported by a nationally representative sample of secondary school teachers in a survey conducted by the Ministry of Education at the end of 2014 (*Encuesta Nacional de Docentes*).

[14]In 2015, the working week was expanded by two (paid) hours, which are meant to be spent performing activities outside the classroom, namely preparing materials for class, assisting students who fall behind, providing orientation to parents, etc. In 2016, an extra two hours were added, reaching a total of 30 working hours a week.

[15]Mizala and Ñopo (2016) examine the patterns of teacher pay in several Latin American countries in an earlier period (1997-2007), and find that teachers in Nicaragua and Peru were the most underpaid relative to their nationals working as professionals or technicians.

[16]Around April 2015, the Ministry of Education launched *Semaforo Escuela*, a program in which trained enumerators make periodic visits to public schools, and register information on teacher, student, and director absenteeism, among other things. Further details are available at http://www.minedu.gob.pe/semaforo-escuela/. In a 2006 study, teacher absenteeism in Peruvian public schools was found to be higher, around 11% (Chaudhury et al., 2006).

1.2.2 The BE Program

In 2013, the Peruvian Ministry of Education launched *Bono Escuela* (BE), a nationwide teacher pay-for-performance program in public schools. The program was first implemented in primary schools, and extended to secondary schools in 2015. Secondary schools, the focus of this paper, were only included in BE starting 2015 because Peru's census standardized tests (*Evaluación Censal de Estudiantes*, henceforth ECE), one of the key indicators used for the BE, were not implemented in secondary schools until 2015.[17] The ECE is an annual low-stakes test designed by the Peruvian Ministry of Education, in which students from different grades in practically all private and public schools are tested on their basic competencies in math and language at the end of the school year.[18] In secondary schools, only 8^{th} graders are tested. The ECE is implemented by the Peruvian National Statistics Institute (INEI), which trains independent enumerators for this task. Since the main goal of the ECE is to track the evolution of student learning throughout the country and help shape educational policies, school average scores are reported to school district governments, schools and parents.

Besides being an informational tool for the Ministry of Education, ECE test scores are one of the metrics used to rank schools and select the BE bonus recipients. Schools not eligible for taking the ECE test in 2015 (only 4% of all public secondary schools) compete for a smaller bonus based on other measures. We focus our analysis solely on public schools taking the ECE. As outlined in Table 1.1, a school's score for the BE is composed of several factors. The score gives 40% of weight to the average math and language grade of 8^{th} graders in the ECE standardized tests.[19] In order to prevent teachers from encouraging ab-

[17] We do not examine the effect of the primary school BE program due to identification issues related to the timing of the program's implementation. The BE was first announced by the president of Peru in July 2014, although the corresponding regulation only came out in October of that year. The 2013 edition was implemented retroactively (i.e., after the 2013 ECE test had been taken), in an attempt to boost the program's credibility. In the case of 2014, it is unclear whether schools knew about the program before taking the ECE in November, since the BE was broadly announced four months before the test, but its regulation only came out one month before.

[18] The ECE was first implemented in 2007 in 2^{nd} grade of primary school, and was extended in the following year to 4^{th} grade in schools with intercultural bilingual education. It was administered in 8^{th} grade for the first time in 2015, and will be extended to 4^{th} graders in all schools in 2016.

[19] In the primary school BE program, the score is also composed of the change in the average

senteeism of low achieving students on the day of the ECE evaluation, schools not complying with a minimum rate of student participation are disqualified from taking part in the BE. In particular, ECE participation must be 80%, 90% and 95% in schools with only one, two or more than two 8^{th} grade classes.[20] Additionally, 35% of weight is given to the entire school's intra-annual retention rate, that is, the proportion of enrolled students still in school at the end of the year. Although dropout rates are non-negligible, most of the dropping out takes places after the school year ends, making retention rates already extremely high before the program was implemented. The average retention rate in the public secondary schools was 99% in 2014, and only 7% of schools had retention rates below 95%. In practice, schools had very little leeway for improving their retention rates, and could thus not compete on the basis of this indicator.[21] An extra 5% of the school's score depends on whether the principal enrolls his students in the Ministry of Education's administrative system (*Sistema de Información de Apoyo a la Gestión de la Institución Educativa*, henceforth SIAGIE) in a timely manner, something which should not affect the incentives of teachers and thus the performance of their students. The remaining 20% of the score depends on an index of school management, composed of teacher attendance, management of school infrastructure, compliance with class hours, as well as measures of pedagogical practices and learning environment. The first three measures are collected by independent evaluators making visits to all public schools, whereas the last two are obtained from questionnaires handed out to 8^{th} grade students during the ECE. All in all, around 80% of a school's score ultimately depends on the performance of 8^{th} grade students in math and language. Consistent with this fact, schools ranked in the top 20% of their BE group according to their average ECE score were 57 percentage points more likely to win the bonus.

The timing of the BE is depicted in Figure 1.2. The school year in Peru starts in March, and ends in December. At the end of the 2014 school year, once the

ECE scores from the previous year, to incentivize schools in the lower end of the distribution. Since 2015 was the first year the ECE was implemented in secondary schools, this could not be replicated.

[20] 91% of public secondary schools complied with this requirement, and the average school only had 1.4 students absent on the day of the exam.

[21] In practice, giving such a large weight to this indicator counteracts any perverse incentives schools might have to improve their test scores by encouraging their weakest students to drop out

implementation of the ECE test in secondary schools was confirmed for the following year, the Minister of Education announced the possibility of extending the BE program to secondary schools as well.[22] The government resolution regulating the 2015 BE came out on the 25^{th} of July, almost four months before the 2015 ECE (carried out in November 17/18), and was accompanied by a diffusion campaign launched by the Ministry of Education informing schools about the BE program. In comparison to other studies, the time frame teachers had to react was relatively short. We elaborate on this issue using the results of our teacher survey in Section 1.7.6.

BE is set up as a collective incentive, such that the principal and every teacher in a school are rewarded if the school scores in the top 20%.[23] To ensure that schools competing against each other are comparable, they are separated into groups by school district, instruction time, and by whether they are urban or rural. There are 395 groups in total, with an average of 34.6 schools per group. Teachers in schools in the top 10% in their group get a bonus of 2000 soles (roughly 605 dollars), whereas those in schools in the top 10%-20% get paid 1500 soles (454 dollars). Every teacher in a winning school gets the exact same bonus, whereas the school principal gets a slightly larger payment (500 extra soles). Since the average teacher receives a monthly salary of 1469 soles, the bonus constitutes either 1 or 1.4 monthly salaries on average. Considering that 20% of schools receive the prize, the average value of the bonus is 24% of a monthly salary, a sizable figure as compared to other studies in the literature.[24,25]

[22] http://larepublica.pe/21-12-2014/jaime-saavedra-el-proceso-para-nombrar-a-8-mil-maestros-se-inicia-en-julio-del-2015 (last accessed August 16, 2016).

[23] Other papers studying collective teacher incentives are Lavy (2002) in Israel, Glewwe et al. (2010) in Kenya, Muralidharan and Sundararaman (2011) in India, Contreras and Rau (2012) in Chile, and Fryer (2013) and Goodman and Turner (2013) in the US.

[24] The average payment is 350 Soles (0.1x2000 + 0.1x1500), which constitutes 24% of the average teachers' monthly salary.

[25] In the teacher incentive program of Muralidharan and Sundararaman (2011), the average bonus was around 35% of a monthly salary; in the experiment run by Glewwe et al. (2010) in Kenya prizes were in-kind, and the average teacher got a bonus worth 12%-21% of a monthly salary. In the Israeli program studied by Lavy (2002) prizes of 10%-40% of the average teacher's monthly salary were awarded to approximately one third of participating teachers, whereas the prizes roughly represented 50% of a monthly salary and were awarded to two thirds of teachers in the New York experiment studied in Fryer (2013) and Goodman and Turner (2013). The incentive implemented in Chile and studied by Contreras and Rau (2012) awarded an average bonus of 10% of a monthly salary.

1.3 Estimation Strategy

We exploit the fact that a school's score for the BE largely depends on the performance of 8^{th} grade students in the math and language ECE test for estimating the causal effect of the teacher incentive on student learning. This feature of the BE results in schools having a much higher incentive to improve the learning of 8^{th} grade students as compared to students from other grades. With this notion in mind, we perform a difference-in-differences estimation comparing the change in achievement of 8^{th} grade public school students with that of 9^{th} grade students attending the same school.[26] In our preferred specification, we use a repeated cross-section of 8^{th} and 9^{th} grade students in public secondary schools eligible for the BE, and run the following regression:

$$\begin{aligned} Internal\, Grade_{ist} = \beta_0 + \beta_1\, 8^{th}\, Grade_{ist} + \beta_2\, 8^{th}\, Grade_{ist} \times Post_t \\ + X_{ist}\,\delta + \gamma_t + \gamma_s + U_{ist} \end{aligned} \tag{1.1}$$

where $Internal\, Grade_{ist}$ is the grade that student *i* in school *s* and year *t* obtained in a particular subject (i.e., the grade assigned to student *i* by his/her teacher at the end of the school year). We run separate regressions using math and language internal grades in our main specification, and also estimate this equation for the average internal grade in all subjects not evaluated in the ECE, to examine whether the BE impacted students' performance in other courses. $8^{th}\, Grade_{ist}$ is a dummy for whether student *i* from school *s* is an 8^{th} grader in year *t*, $Post_t$ is a dummy taking the value of one in the year 2015 and zero in 2013-2014, X_{ist} is a set of individual controls (gender, if Spanish is the student's native tongue, if the student was retained in the previous year, has a disability, and whether the parents are alive and living in the same household), and γ_t and γ_s are year and school fixed effects. We run regressions for the period 2013-2015, i.e., two years before the BE, and the year in which it took place. Our estimation thus compares students in

[26]We use 9^{th} grade as our comparison group and not 7^{th} grade, for example, because the program might have an impact on teachers in the latter grades, given that their students will be taking the ECE standardized test in 2016.

8^{th} and 9^{th} grade, within the same school, before and after the BE was introduced. Including school fixed effects allow us to restrict our comparison to students facing the same educational environment, but differing in their exposure to the BE.[27] U_{ist} are all the unobserved determinants of achievement for student i in school s and year t, such as ability, motivation, household income, and home environment, to name a few. We allow for our errors to be correlated within school by clustering our standard errors at the school level. We express grades as a z-score, standardizing them by subject and year, so that our coefficient of interest (β_2) can be interpreted as the standard deviation (SD) change in internal grades associated with the incentive. In the case of non-incentivized courses, we first calculate the z-score for each course, and then take the average. As a robustness check, we also standardize internal grades for each subject by school and year.

Unlike other studies on teacher pay-for-performance, our outcome variable is the grade assigned to students by their teachers at the end of the school year (what we refer to as internal grades), and not their standardized test results. Given that teachers' pay under the BE is tied to performance in the ECE, an identification strategy relying on standardized test scores as an outcome cannot disentangle whether improvements in students' performance are the consequence of increased student learning or the results of short-term strategies fostering high test scores (Neal, 2011). Having internal grades as our outcome has the advantage of capturing students' performance without directly influencing teachers' probability of obtaining the bonus. While it would still be interesting to study the impact of the BE on students' ECE test scores, we cannot do so because the ECE test was applied in secondary schools for the first time in 2015, and there is no group of students serving as an appropriate comparison. From the perspective of students, internal grades play a very important role, directly affecting whether they pass the school year, take summer remedial courses or are retained. Importantly for

[27] Given that 8^{th} grade students in private schools take the ECE but these institutions are not eligible for the BE, we could also run a differences-in-differences regression comparing the change of internal grades of 8^{th} grade students from public and private schools, similar to what Contreras and Rau (2012) do for the case of Chile. However, as shown in Appendix Table 1.A1, public school students were already improving relatively faster than their private school counterparts in the year prior to the BE (i.e., the *Public* $\times$ *2014* coefficient is statistically significant). Since there are other things that could be changing across the public-private spectrum in 2015 that we cannot control for, we discard this estimation strategy.

our identification, teachers' grading tactics should not be influenced by the BE incentive. Since the bonus from the BE is tied to ECE test results, and not internal grades, teachers' stakes in their students' internal grades are not directly modified by the incentive program.[28] Although it is mandatory for schools to report students' internal grades to the Ministry of Education, these grades have absolutely no bearing on whether the school obtains the bonus. As long as internal grades present enough variability, we would expect them to reflect changes in learning. We provide supporting evidence of the fact that within-schools, internal grades are correlated with standardized measures of learning, and show that grading on a curve is uncommon in Peruvian secondary schools in Section 1.6.2.

Our main identifying assumption is that in the absence of the teacher incentive, the performance of 8^{th} and 9^{th} grade students attending the same school would have evolved in an equivalent way between 2014 and 2015. A necessary condition for giving a causal interpretation to β_2 is that 8^{th} and 9^{th} grade students follow parallel trends before the implementation of the BE. An inspection of the raw means in Figure 1.1 shows that grades of 8^{th} and 9^{th} grade students appear to be on parallel trends in both math and language before the program was implemented. We provide formal evidence for the parallel trends assumption in Section 1.6.1.

Identifying a causal effect also requires that the performance of 9^{th} grade students, our comparison group, is unaffected or hardly affected by the teacher incentive program (i.e., that there are no spillovers). Importantly for our identification, schools do not have much room to compete on the basis of indicators other than the 8^{th} graders' standardized test scores, leading to a practically null correspondence between 9^{th} grade students' learning and a school's BE score. As explained in Section 1.2.2, around 80% of a school's score ultimately depends on the performance of 8^{th} grade students. This implies that, if any, a very small portion of the school's score could be improved if 9^{th} grade teachers exerted more effort. It is important to bear in mind, however, that since 83% of 8^{th} grade teachers also instruct 9^{th} grade, an increase in effort while teaching 8^{th} graders could potentially

[28]Our teacher survey inquires, among other things, about whether teachers changed the difficulty of their classes in 2015 as a result of BE. As shown in Table 1.10, teachers were equally likey to report that they decreased the difficulty of their classes when teaching students from 8^{th} grade, as compared to students from other grades, and only 5 percentage points more likely to report that they increased the difficulty of their classes.

spill over to students in our comparison group and bias our estimates downwards. We show that this is not a concern by exploring the impact of the teacher incentive in schools with a low overlap between 8^{th} and 9^{th} grade teachers in Section 1.6.2. On the other hand, the fact that the probability of obtaining the bonus hinges largely on the performance of 8^{th} grade students might lead the school to redirect its resources towards these grades, negatively impacting the internal grades of students in our comparison group. We discuss this in further detail in Section 1.6.3, and show that this issue is not a concern in our setting.

1.4 Data and Descriptive Statistics

1.4.1 Administrative Data

Our empirical exercise relies on a rich administrative database collected by the Peruvian Ministry of Education in 2013-2015, derived from its SIAGIE system. Coverage is basically universal, reaching 99.7% of public schools. Schools must enroll their students into the SIAGIE system at the start of the school year, and input the final grades of their entire student body once the academic year concludes. Grading is done on a 0-20 scale, and students need to obtain at least 11 to pass a given subject. Besides students' grades, this database also has information on characteristics such as age, gender, native tongue, parents' education, if they live with their parents, etc. Student identifiers permit tracking individual students across years. The SIAGIE also contains information on the grade and classroom that student are assigned to, the teachers who teach each grade and group, and some basic teacher characteristics such as age and gender. In 2015, there were 8,654 public secondary schools in Peru, of which 8,092 were eligible for participating in the ECE. Schools must have at least five 8^{th} grade students in order to be eligible for taking the test. Our SIAGIE database covers 8,059 of these schools.

Table 1.2 presents some characteristics of the 8^{th} and 9^{th} grade students attending public secondary schools in 2013-2015. We observe that the mean final grade in math is 12.27 and 12.32 (out of 20) in 8^{th} and 9^{th} grades, and 84% and 85% of students pass this course. Students perform slightly better in language, where the mean final grade for the 8^{th} and 9^{th} grade students is 12.67 and 12.64, and 89% and 90% of them pass the course. Mean grades in other courses exceed those of math and language by almost one point, and almost all (93% and 94%) students pass these courses. Half of the students are male, almost all of them are natives, and 83-84% of them have Spanish as a native tongue. Only 6% and 4% of 8^{th} and 9^{th} graders were retained in the same grade the year before. Although it is not necessary for our identification that 8^{th} and 9^{th} grade students are balanced in terms of observables, they do appear to be very similar. In addition, Table 1.2 shows some characteristics of the 8,059 public secondary schools in our sample. Less than half (40%) of the public secondary schools are located in rural areas. Each school has, on average, two classes per grade, and there are around 19 students per teacher in the average class. We also observe that each school has, on average, roughly 11 teachers teaching 8^{th} and 9^{th} grade, with 83% of teachers in 8^{th} (9^{th}) grade also teaching 9^{th} (8^{th}) grade. Teachers are almost 42 years old on average, and 60% of them are male.

1.4.2 Survey Data

We complement our main empirical analysis with the results of an online survey we conducted with the assistance of the Ministry of Education. According to our teacher database, there were 123,669 public secondary school teachers in 2015. The Ministry of Education has the email address of 36,283 of them (30%), all of which received a survey email from the Ministry in October 2016, a few weeks before the winners of BE were announced. As in the past editions of BE in primary schools, the bonus winners were announced at the end of the following school year (in November 2016). Teachers were asked what grades and subjects they taught in 2015, their knowledge about the BE and its rules at that time, and their opinion about the size of the bonus. We also inquired about changes in their pedagogical practices while teaching students from different grades, and about administrative changes in the school they were working for in 2015. Finally, we tried to elucidate teachers' perception about their school's ranking and its probability of winning, and asked teachers for their opinion about students' motivation in the standardized test tied to the BE.

The survey was anonymous, and teachers were told that its purpose was to collect information about teachers perceptions and opinions about the BE program. Since the survey was framed in the context of BE, and sent by the Ministry of Education, respondents might be subject to social desirability bias (i.e., over-reporting of good behavior associated with the objectives of BE). To try and maximize the response rate, and due to restrictions imposed by the Ministry, we did not ask questions about teacher characteristics or identify the school they worked for, and thus we cannot compare survey respondents to non-respondents. We received a response from 3,406 teachers (9.4% response rate), roughly 2.8% of all public secondary school teachers. Given the potential bias in teachers' responses and our selected sample, the results from this survey must be taken with caution.

1.5 Results

The teacher incentive program had no effect on 8^{th} grade students' math and language internal grades, as shown in columns (1) and (4) of Table 1.3. Our coefficient of interest (the interaction of the 8^{th} *Grade* and *Post* dummies) is robust to the inclusion of school fixed effects (columns 2 and 5) and individual controls (columns 3 and 6), with the latter being our preferred specification.[29,30] Our coefficients are precisely estimated zeros, allowing us to reject positive effects larger than 0.008 SD in math, and 0.017 in language, well below the treatment effects found in the existing literature. In the teacher incentive program studied by Muralidharan and Sundararaman (2011) in India, average math and language test scores increased by 0.15 SD after one year, whereas Contreras and Rau (2012) find that a teacher incentive program in Chile had positive and large effects on language and math test scores of 0.14-0.25 SD. While the incentive scheme evaluated by Glewwe et al. (2010) in Kenya led to a 0.14 SD increase in test scores in tests linked to the incentive, the authors found no impact on the outcome of non-incentivized evaluations, consistent with our findings.

Since there are 395 distinct groups in which schools compete for the BE bonus,

[29] In our baseline regressions we standardize internal grades by subject-year, but the results are quantitatively similar if we standardizing each subject by school and year.

[30] Parents' education is missing for 12% of students, so we do not control for this in our baseline regressions. However, attrition is not differential across grades, and our results are robust to controlling for this.

i.e., 395 different tournaments, we also evaluate the average effect of the teacher incentive in every competition. Figure 1.3 displays the 8^{th} *Grade* x *Post* coefficients (and its 95% confidence interval) for math and language in each of these 395 tournaments. In the vast majority of these groups, the teacher incentive had a zero average effect on student achievement. The coefficients for math and language are positive and statistically significant at the 5% level in only 4% and 6% of the BE groups,[31] providing further evidence of the BE's null average effect on student achievement.

As in most comparable studies, teacher bonuses under the BE are tied to students' performance in just two subjects (math and language). However, teacher incentives might also have an impact on student learning in other courses. The sign of this impact is theoretically unclear. On the one hand, schools could be tempted to devote more resources towards math and language at the expense of other subjects (e.g., augmenting instruction time), negatively impacting learning in the latter. On the other hand, 8^{th} grade teachers in all subjects, not just math and language, might exert more effort knowing that their school's score largely rests on the performance of these students. Additionally, due to complementarities, if learning were higher in math and language, student achievement in incentivized subjects might increase indirectly (Muralidharan and Sundararaman, 2011).[32] We do find a positive and small but significant effect of 0.011 SD on grades in non-incentivized courses, as shown in Table 1.4. Appendix Table 1.A2 breaks the results down by each of the nine non-incentivized courses; we observe positive effects ranging between 0.014 SD and 0.017 SD in three cases (social studies, human relations, and religion). Although significant, the observed effect is very small, and well below the spillover effects found in other papers.[33] Furthermore, these results should be taken with caution because, as further discussed in Section 1.6.1, there is a divergence in the trend in non-incentivized courses in the year

[31]Furthermore, in only 6 out of the 395 BE groups this holds simultaneously for math and language.

[32]Unlike studies carried out in primary school, math and language teachers are not responsible for teaching other subjects in secondary school. Thus, if there were any positive spillovers to other courses, they would be indirect.

[33]Muralidharan and Sundararaman (2011) find that teacher incentives targeted towards math and language standardized tests had an effect of 0.11 and 0.14 SD in science and social studies after only one year, an effect 10 to 13 times larger than the one we find.

before the program was implemented.

1.5.1 Heterogeneous Effects

In a tournament such as the BE, if teachers are risk neutral, have symmetric information, and if students in all schools have the same ability (i..e, if all schools have the same ex-ante probability of winning), all teachers should exert the same effort as a result of the incentive, and who gets awarded the bonus should be random (Lazear and Rosen, 1981). However, if schools differ in their probability of winning, the incentive might not have the same power across the board. For example, teachers in schools in which students' pre-program levels of achievement are very far from the top 20% could be discouraged from exerting extra effort, and schools which are almost guaranteed to win might not respond to the incentive. This concern is partly mitigated in our setting by the fact that schools are grouped according to characteristics which are likely correlated with their students' performance, such as their number of hours of instruction, whether they are urban or rural and their school district. Nevertheless, important differences between schools in the same group might still remain, possibly affecting the reach of BE. This notion is brought forward in the Chilean study of Contreras and Rau (2012), where the authors find that the teacher incentive only had a positive impact on schools above the 65^{th} percentile in the distribution of pre-program score (the program awarded a bonus to schools in the top 25% within their group). The fact that the ECE was implemented in secondary schools for the first time in 2015 provides a limitation for performing this analysis in our context, since we cannot accurately determine a school's pre-tournament probability of winning. As a second best, we proxy a school's likelihood of winning using its relative ranking within its BE group in terms of the socioeconomic status (SES) of its students. We construct an average measure of the SES of 8^{th} graders in 2015 by considering whether their first language is Spanish, and whether their parents have more than a primary school degree.[34] We then rank schools within their BE group according to this measure, and fully interact our baseline regression with 20 dummies indicat-

[34]For each 8^{th} grader in 2015, we add three dummy variables: whether his first language is Spanish, and dummies for whether his mother and father have more than a primary school education. We then calculate the average index for each school.

ing the percentile in the within BE group distribution that each school belongs to. As depicted in Figure 1.4, the estimates for all of these percentiles are very small in both math and language, and most of them are not statistically significant.[35] Having said this, it is highly unlikely that schools knew their relative standing in their BE group and could anticipate the likelihood of winning. We discuss this in Section 1.7.5.

From a theoretical perspective, the strength of the incentive might be decreasing in the number of 8^{th} grade teachers and/or students, since the marginal impact of a teachers' effort on its school's score decreases when there are more teachers and students reached by the incentive, and teachers' ability to monitor each other also diminishes. For instance, Imberman and Lovenheim (2015) find that the effect of a group-based teacher incentive program in Houston is much stronger when teachers are responsible for teaching a higher share of students. Since our teacher database does not have information on the subject that each teacher is responsible for, we do not know how many incentivized teachers each school has; as a second best, we use the number of 8^{th} grade classes in 2015 as a proxy. We do not find any significant interaction of the BE incentive with enrollment or number of groups per grade, as seen in columns 6 and 7 of Table 1.5. Finally, we do not find any effects by whether the school ir urban or rural, as shown in column 8. As with any heterogeneity analysis, it is important to take the results with caution, since characteristics such as enrollment and urbanicity are not randomly assigned, and could be proxying for something else. Ideally, we would also be able to test for heterogeneous effects across teacher characteristics. Unfortunately, although we know who the teachers are for each class, we do not know which of the teachers teach math and language.

Following other papers in the literature, we also test for heterogeneous effects across gender, by whether students' first language is Spanish, and by their parents' educational attainment. The latter variable is an index from 0 to 2, taking a value of 0 if both parents have a primary school degree or less, 1 if one parent has more than primary schooling, and 2 if both do. Parents' education and students' native

[35] We also perform this exercise ranking schools instead by an index measuring the quality of their infrastructure, another proxy of their probability of winning, and find no discernible patterns either (results upon request).

tongue are proxies for socioeconomic status in Peru. As displayed in column 1 of Table 1.5, and consistent with the finding in Muralidharan and Sundararaman (2011) and Behrman et al. (2015), we do not find any heterogeneity by gender. Neither do we find heterogeneous effects by socioeconomic status, proxied by native language and parents' education (columns 2 and 3). The literature is mixed on this particular issue, since Muralidharan and Sundararaman (2011) observe that students from more affluent families have a stronger response to the teacher incentive program, whereas Lavy (2002) finds that it is students with poor socioeconomic backgrounds that benefit more from it. However, the program evaluated in the latter was designed so as to encourage teachers to focus on weak students.

Considering that teachers might focus on certain students, and student responsiveness might vary according to prior achievement, we also test for heterogeneity across measures of students' past performance, namely whether the student was retained in the previous year and by the student's lagged internal grade in the same subject (standardized by school, grade and year).[36] Pay-for-performance programs in which bonus payments depend on whether students attain a certain threshold, such as passing an exam, create incentives for teachers to focus on students close to this cutoff (e.g. Lavy, 2009 and Neal and Schanzenbach, 2010). On the contrary, if obtaining the bonus depends on the average score, such as in the BE program under analysis, teachers will find it optimal to target students most responsive to any increased teacher effort. If the function mapping teacher effort into test score gains is concave (convex) in past performance, teachers would react by focusing more intensely on the weaker (stronger) students (Muralidharan and Sundararaman, 2011). However, as shown in columns 4 and 5, we do not find any heterogeneity according to students' past performance.[37] These results are consistent with the findings of Behrman et al. (2015).

[36]Lagged grades are only available for students in 2014 and 2015, since our database only has student identifiers which can be linked across years starting 2013. Importantly, if we restrict our sample to this period, results on average treatment effects do not change.

[37]We also perform this estimation by grouping students into quintiles and terciles of the distribution of lagged grades in their same school, grade and year. The results are unchanged, as reported in Table 1.A3.

1.6 Testing the Validity of the Identification Strategy

This section provides further evidence on the validity of our difference-in-differences estimation strategy. We provide formal evidence in support of the parallel trends assumption, and demonstrate that internal grades are broadly correlated with ECE test scores, and vary considerably within schools. Furthermore, we corroborate that our null effects are not driven by positive spillovers to our comparison school, and show that schools did not change the way in which they assigned teachers across grades as a result of the teacher incentive program.

1.6.1 Parallel Trends

To test whether there is a divergence in the trends of 8^{th} and 9^{th} grade students in 2014, we add an interaction between the 8^{th} grade dummy and an indicator for 2014 to our baseline specification. Reassuringly, the coefficients for the pre-treatment difference-in-differences are precisely estimated zeroes for both math and language, as shown in Table 1.6. In the case of non-incentivized courses, however, there is a relative increase in 8^{th} graders' internal grades in 2014. Although the magnitude of this change is small (0.010 SD), it is similar in magnitude to the estimated impacts for 2015. Hence, the results using non-incentivized courses as an outcome should be taken with caution.

1.6.2 Internal Grades Reflect Learning

Unlike other studies on teacher pay-for-performance, we measure learning using students' internal grades instead of their standardized test results.[38] As discussed in Section 1.3, internal grades have the advantage of capturing student achievement without directly influencing teachers' probability of obtaining the bonus. However, internal grades are subjectively assigned by teachers, and are not awarded using a uniform criterion as standardized tests are. Since each school

[38] In a recent study, Chong et al. (2016) also use internal grades to measure student achievement in rural Peru.

might have its own grading standards, making differences in internal grades not necessarily reflective of differences in learning across schools, we restrict our comparison to students from the same school to control for school-specific grading standards.[39] What is crucial for identifying a causal effect is that internal grades capture changes in learning across different grades within the same school. That is, if 8^{th} grade students in a particular school are learning more as a result of the teacher incentive, the relative internal grades of 8^{th} graders in that school should rise. We face two potential threats in this regard. Firstly, teachers might not award internal grades in a systematic way. This doubt is raised by the findings of a few papers comparing grading standards in blind versus non-blind examinations. While some studies find evidence of discrimination in grading based on students' gender (Lavy, 2008), ethnicity (Botelho et al., 2015; Burgess and Greaves, 2013), and caste (Hanna and Linden, 2012), others find no such disparities (Newstead and Dennis, 1990; Baird, 1998; Van Ewijk, 2011). Consistent with the latter, we show that student characteristics correlate with internal grades and with standardized test scores in a consistent manner within the schools in our sample, alleviating this concern. A second threat to our identification is that if teachers grade on a relative basis (e.g., the worst 10% always fails, or the top 10% always gets the highest grade), we might not be able to detect overall changes in student learning using internal grades. It turns out, however, that there is substantial variation in the distribution of grades across classes and years in the same school.

Considering that our identification requires that internal grades reflect within-school differences in learning, standardized test scores and internal grades should broadly follow the same patterns when comparing students from the same school. Unfortunately, ECE test scores are disclosed at the school level, meaning that for every secondary school taking the ECE in 2015, we only observe the mean score in math and language, as well as the fraction of 8^{th} graders with very low, low, medium and high performance. Given our data limitations, we examine the cross-

[39] Although it would be preferable to include teacher fixed effects to control for teachers' grading standards, we only know the grades and classes teachers are assigned to, but not the subject that they teach. We cannot identify who the teacher handing out the grades for each subject is, and therefore cannot include teacher fixed effects in our estimation. However, since teachers are not systematically changing across 8^{th} and 9^{th} grades, as shown in Section 1.6.2 below, unobserved teacher characteristics are unlikely to bias our estimates.

sectional correlation between average ECE scores and the average internal grades of 8^{th} graders in 2015, for the public secondary schools in our sample.[40] We also explore the correlation between the fraction of students who fail math and language according to their internal grades, and the fraction of low performing students in the ECE. To facilitate the interpretation of the coefficients, we express average internal grades and average ECE scores as a z-score, and control for school district fixed effects, school characteristics and the average characteristics of students from each school. As shown in Appendix Table 1.A4, average ECE scores and internal grades are positively and significantly correlated, although their correspondence is relatively weak. In particular, a 1 standard deviation increase in average math (language) internal grades is associated with an increase in average ECE scores of 0.116 (0.103) SD. Moreover, a 1 percentage point increase in the share of students failing math (language) according to their internal grades corresponds to a 0.071 (0.091) percentage point rise in the proportion of students with the lowest attainment in the ECE.[41,42]

Having said this, it is hard to establish whether internal grades reflect learning by just comparing the aggregate cross-sectional correlation of these and ECE grades. For one, internal grades might capture a related but different dimension of learning than standardized test scores. Additionally, since internal grades are likely to depend on school grading standards, it is unclear that they can be compared across schools.[43] While the disclosure of ECE test scores does not allow us

[40]Our sample for this analysis (8,010 schools) is slightly smaller than our baseline sample of 8,059 schools because a few schools with were eligible to take the ECE (and were thus eligible to participate in the BE) ended up not taking the test, or were faced with problems during its implementation.

[41]In 2015, 55% and 43% of 8^{th} graders in the average school were ranked in the lowest category according to their ECE scores, whereas the average school only had 13% and 9% of their 8^{th} graders failing math and language, respectively. These two categorizations are only broadly comparable, and these results must thus be taken with caution.

[42]We also examine whether school and average student characteristics explain internal and ECE grades in a similar manner, by separately regressing schools' 2015 average ECE and internal grades against a series of controls. As displayed in Appendix Table 1.A5, the same broad patterns hold for both types of grades in math and language. Schools in which a high proportion of students have parents with more than a primary school degree do better as reflected by both ECE and internal grades. The same holds for schools with longer school days, and schools in which a high proportion of students have Spanish as their first language, Furthermore, schools in which a higher share of 8^{th} graders were retained the year before do worse according to both measures.

[43]If good schools set harsh grading standards, and low quality schools are lenient in their

to identify students' individual performance, the Peruvian Ministry of Education provides an anonymized database with individual ECE test scores, gender, an index of socioeconomic status (constructed using parents' education, and household assets and characteristics), and anonymized school identifiers. As shown in Panel A of Table 1.A6, students are more likely to obtain a higher ECE test score in math and language if they are male and have a high socioeconomic status, as compared to other students from the same school. An analogous regression with 8^{th} grade students' individual internal grades as the dependent variable (Panel B of Table 1.A6) shows that the within-school correlation between student achievement and gender and socioeconomic status is qualitatively similar. Despite the fact that internal grades and standardized test scores are prone to measure learning differently, and that students have different stakes in each of these outcomes, these two measures seem to relate in a consistent manner when comparing students from the same school.

Having established that internal grades are correlated with standardized test scores, we now provide evidence of the fact that grading on a curve is uncommon in Peruvian secondary schools. If teachers were assigning grades on a relative basis, we would expect two different classes in the same school, grade and year to have a very similar grade distribution. Our database on teachers shows that on average, 8^{th} grade teachers from schools with only two classes teach in 92% of them, meaning that the teachers handing out the grades are practically the same across classes. We restrict our sample to 8^{th} graders in schools with just two 8^{th} grade groups in 2014 (accounting for 17% of our schools), and test whether math and language internal grades have a different mean and standard deviation across both classes belonging to the same school. With a significance level of 10%, in 23% and 32% of cases we reject the null hypothesis of equal means across both groups in math and language, respectively. The average difference in means across groups is 0.66 and 0.77 in math and language, roughly one third of a standard deviation. An F-test for the equality of variances shows that in 23% and 21% of our schools, we can reject the null hypothesis that the distribution of math

grading, for example, differences in the average internal grades of these two types of schools will not convey any information on their differences in student achievement.

and language grades has the same standard deviation.[44] The difference in means and standard deviations and their corresponding p-values are depicted in Figure 1.A1. All in all, this evidence points to the fact that grading on a curve is not the norm in Peruvian high schools.

1.6.3 No Spillovers to 9^{th} Grade Students

A crucial condition for identifying the causal impact of BE is that the performance of 9^{th} grade students, our comparison group, should be unaffected by the program. One possible concern is that the incentives introduced by BE could lead schools to change the assignment of teachers across 8^{th} and 9^{th} grade in 2015. If schools assigned the best teachers to 8^{th} grade in 2015, for instance, our estimates would be upward biased. However, since we find that BE had a null effect on student achievement, this is less of a concern. Furthermore, since the program was only announced after the school year had started, as illustrated in Figure 1.2, it would have been hard for schools to shift teachers around. We still perform some tests to learn whether schools teacher characteristics changed differently across grades in the year in which BE was introduced. Given that we cannot identify which teachers are responsible for instructing math and language, and do not have an objective measure of teacher quality, it is hard to test if BE brought about changes in the average quality of teachers across grades. However, we do observe the school, grades and classes to which teachers are assigned to in 2013-2015, and have some observable teacher characteristics which might be correlated with their performance. We use this information to test for differential changes in 2015 in the average characteristics of 8^{th} and 9^{th} grade teachers from the same school. As shown in Table 1.8, we do not find any differential change in the average age and gender of teachers in 2015. Neither do we observe a significant change in the proportion of teachers who are new to the school or new to that particular grade and school, or in the average number of courses taught by teachers in that grade. We do observe a significant decrease in the average number of secondary schools in which 8^{th} grade teachers are working. Although this might mean that 8^{th} grade teachers were less time constrained in 2015, this only represents a 0.3%

[44] The average difference in standard deviations is 0.44 in math and 0.41 in language.

drop from the mean. Consequently, there is no strong evidence of changes in teacher composition across 8^{th} and 9^{th} grade classes belonging to the same school in 2015.

A bigger concern given our findings is that BE improved teacher behavior overall, instead of impact teaching to 8^{th} graders differently. As we explain in Section 1.2.2, in practice around 80% of a school's score depends on the performance of 8^{th} grade students in the ECE standardized tests. Thus, a very small portion of the school's score could be improved if 9^{th} grade teachers exerted more effort. However, since 83% of 8^{th} grade teachers also instruct 9^{th} grade, any increase in effort while teaching 8^{th} graders could spill over to students in our comparison group, biasing our estimation downwards. Alleviating this concern, we find that the effects are also null in schools in which a low share of 8^{th} grade teachers also instructs 9^{th} graders (columns 1, 2, 4 and 5 of Table 1.7).[45] If anything, there is a significant (though very small) positive effect in math grades in schools in which most teachers instruct both 8^{th} and 9^{th} graders (column 3).

1.7 Why Didn't Student Learning Increase?

Having established that student learning did not increase as a result of the teacher incentive program, this section discusses and provides suggestive evidence on a series of potential explanations for why the program had a null effect.

1.7.1 Teachers Did Not Know About the Program or Did Not Understand It

An explanation for the null effects we find is that schools simply did not hear about the BE, or did not understand the formula by which scores were calculated. We argue that this is at best a partial explanation. In 2015, along with the launch of the 2015 edition of the BE program, the Ministry of Education headed a diffusion campaign, making it likely that secondary school teachers were informed about

[45] Even though what matters is the overlap of math and language teachers, we do not have information on the subjects taught by each teacher, and are thus restricted to perform this analysis using the average overlap of all teachers across 8^{th} and 9^{th} grade.

the program. Furthermore, the fact that the principal and every teacher get paid if their school wins generates strong incentives for people working in the same institution to inform each other about the BE. In our teacher survey, 64% of those who taught math or language in 8^{th} grade in 2015 reported that they knew about the program's existence during the 2015 academic year. When asked about how they heard about BE, 57% answered that they found out through the Ministry of Education or the school district authorities, 30% answered that they got the information from the news, and 35% from the school principal or other coworkers (they could select more than one option).

Although we only evaluate the effect of BE in its first year, the system by which schools were scored under the BE was not overly complex.[46] It should have been relatively clear from a teacher's perspective that the main component of his/her school's score is the average performance of 8^{th} graders in the ECE standardized tests. This stems from the fact that performance in the ECE test was the main component of schools' scores in the two previous rounds of the BE in primary schools. The BE program had already been going on for two editions in every public primary school in the country, and the experience of primary schools with the BE was salient in the national news.[47] This is broadly confirmed by our survey, in which 64% of math or language 8^{th} grade teachers who knew about BE in 2015 answered that ECE test scores were the most important or second most important component of the BE score.

Almost half of the schools in our sample share the building with a primary school that participated in the BE before, and 13% of them operate in the same building as primary school BE winner. Even though the salience of the secondary school BE was probably higher in these cases, we do not find any effects on math and language test scores in either of these groups of schools, as shown in columns (3), (4), (8) and (9) of Table 1.9. Thus, it is unlikely that the BE had no impact because of schools' lack of awareness of its existence or its rules.

[46]Other studies on teaching incentives with similar formulas for assigning the bonus (Lavy, 2002 and Contreras and Rau, 2012) find positive and significant effects on student learning.

[47]For instance, http://larepublica.pe/23-10-2014/maestros-tendran-bono-de-hasta-3-mil-soles-por-buen-desempeno and http://www.andina.com.pe/agencia/noticia-bono-hasta-s-3000-buen-desempeno-docente-se-pagara-noviembre-528482.aspx

1.7.2 The Incentive Was Too Small

The prize that teachers could receive under the BE is in the range of bonuses granted in other studies finding positive effects. As described in Section 1.2, the BE bonus corresponds to either 1 or 1.4 monthly salaries of the average teacher, and was awarded to teachers in 20% of schools. The average bonus represents approximately 24% of a monthly salary, making this incentive sizable in comparison to that of other studies, in which the average value of the prize ranges between 3% and 35% of a monthly salary.[48] In the subsample of 8^{th} grade math or language teachers who responded our survey, 42% correctly identified the bonus amount or thought that it was larger, 20% did not know the exact bonus amount, 2% thought that it was smaller, and 36% did not know about the BE in 2015. However, when we asked their opinion on the size of the prize, only 30% of those who knew about the program thought that the prize was adequate or large. This may have to do with the fact that the survey was coming from the Ministry of Education, and many teachers took this as an opportunity to complain about their low salaries.[49]

If the bonus were not large enough to incentivize the average teacher, we would perhaps find a positive effect in schools in which teachers' pay is relatively low. However, as shown in columns (1) and (6) of Table 1.9, we do not find any heterogeneity by teachers' average salary in 2015.[50] Although we cannot exclude that the incentive scheme would have worked with a larger bonus, there is no evidence that the size of the incentive is the reason why the program had no distinguishable effect on students' math and language grades.

[48] The average bonus obtained by Indian teachers in Muralidharan and Sundararaman (2011) is around 35% of a monthly salary, whereas bonuses in the experiment run by Glewwe et al. (2010) in Kenya have an average value ranging between 12% and 21% of a teachers' monthly wage. The incentive implemented in Chile and studied by Contreras and Rau (2012) paid teachers 10% of a monthly salary on average. Finally, the Israeli program studied by Lavy (2002) awards prizes of 10%-40% of an average teacher's monthly salary to approximately one third of participating teachers.

[49] In the open-ended part of this question, many teachers answered that their salaries are insufficient. Furthermore, quite a few teachers answered the survey email with complaints about their working conditions.

[50] We calculate the average salary of teachers in every secondary school from the number of contract teachers and civil servant teachers in each pay scale, as reported in the 2015 school census. Since the school census does not provide disaggregated data by grade, we take the school average.

1.7.3 Group Incentives Do Not Work

When incentives are collective, the mapping of a teachers' actions on his/her probability of obtaining the bonus is weaker when the number of teachers reached by the incentive is larger, raising the likelihood of free-riding (Holmstrom, 1982), and thus lowering the incentive's power in promoting higher teacher effort. While collective incentives have the potential of inducing higher cooperation and monitoring among teachers (Kandel and Lazear, 1992; Kandori, 1992), this might be harder to achieve when the number of incentivized teachers is very large. Although we do not know the fraction of 8^{th} grade students that each math and language teacher instructs (we do not have information on the subject taught by teachers), we do know the number of 8^{th} grade classes that each school has in 2015. In 2013-2015, the average secondary school in our sample had only two groups of 8^{th} graders. Since there is at most one math and language teacher per group, the average school has no more than four incentivized (i.e., math and language) teachers, a figure comparable to the number of incentivized teachers in other papers in the literature finding positive effects when teacher incentives are collective. The average school in Muralidharan and Sundararaman (2011) and Glewwe et al. (2010) has three and six incentivized teachers, for example.

As shown in column (7) of Table 1.5, we don't find any differential effects by the number of 8^{th} grade groups. If we break the results down even more, as shown in column (2) of Table 1.9, we do find that the BE had a small but significant positive effect in the math grades of students in schools with only one class per grade (accounting for 21% of students in 61% of schools). More specifically, the teacher incentive increases math grades by 0.019 SD,[51] although these effects are much smaller than those found in the other studies in the literature. Thus, the fact that the incentive faced by teachers under the BE is collective does not seem to be one of the main reasons why the program had no effect, although it might have some bite.

[51] The sum of the 8^{th} *Grade* $\times$ *Post* and 8^{th} *Grade* $\times$ *Post* $\times$ *One Class* coefficients yields a total effect of 0.019 SD, with a p-value of 0.008.

1.7.4 Teachers Only Focused on Improving Standardized Test Scores

As discussed in Section 1.3, teacher incentive programs might not result in higher learning if teachers focus their efforts on short-term strategies aimed solely at increasing standardized test scores. Teachers might have reacted to the incentive by targeting topics likely to appear in the ECE, coaching students on test-taking strategies, or even cheating. Since 2015 was the first year in which students took the ECE, and there is no appropriate control group (every public school in which 8^{th} graders participated in the ECE is also eligible for the BE), we cannot identify whether ECE test scores increased as a result of the teacher incentive program. Thus, we cannot initially rule out this hypothesis. However, there are reasons why we believe that teachers could not engage in this type of behavior. Firstly, independent officials, trained and working directly for the Peruvian National Statistics Institute were in charge of the implementation of the ECE. Teachers were not allowed to be in the room at any moment during the exam and were not responsible for its correction. Thus, it is very unlikely that schools could cheat.[52] Secondly, because the ECE exam is designed to capture a wide range of skills,[53] teachers could hardly influence this outcome by narrowing their instructional focus on certain topics. Thirdly, due to the fact that the ECE was implemented for the first time in secondary schools in 2015, secondary school teachers did not have previous experience with this type of standardized tests and, consequently, could hardly predict the content or the specific format of the exam. As the content of the standardized exam was not predictable, coaching or narrow teaching are less of a concern in this setting (Neal, 2011).

Having said this, our online survey inquired about whether teachers changed their pedagogical practices in 2015 as a result of BE, and separately asks about their pedagogical changes while teaching 8^{th} grade as opposed to all other grades. Table 1.10 reports the results of this question for all math and language teachers

[52]Since students had no personal stake in this exam, there were no incentive to cheat on their part.

[53]Details on the design of the ECE are reported by the Ministry of Education in *Reporte Técnico de la Evaluación Censal de Estudiantes (ECE 2015)*, available at http://umc.minedu.gob.pe/wp-content/uploads/2016/07/Reporte-Tecnico-ECE-2015.pdf.

taking the survey who reported that they knew about BE in 2015 (those who did not know where not asked this question). These results must be taken with caution, since it is probable that there was some bias in reporting given the framing of the survey in terms of the BE program.[54] As can be seen in Panel A, 8^{th} grade teachers are 5 percentage points more likely to report that they improved their attendance, and 10 percentage points more likely to report that they gave their students more homework, evaluated them more often and/or gave extra tutoring sessions, as compared to math and language teachers from other grades. There are statistically significant differences as well in how often they report that they paid attention to the weakest students (5 percentage points), increased the difficulty of their classes (6 points), and increased the frequency of multiple choice examinations (9 percentage points). The same patterns hold when we restrict the analysis to teachers that taught math or language in 8^{th} grade and other grades, as seen in Panel B. While some of these self-reported differences in teacher behavior are consistent with teaching-to-the-test (e.g., increasing the frequency of multiple choice evaluations), if teachers were in fact improving their attendance or paying more attention to the weakest students, student achievement in terms of internal grades should have increased, and it did not.

1.7.5 Teachers Were Unfamiliar with the Standardized Test and Students Had No Stakes in It

Given that 2015 was the first year in which the ECE test was implemented in secondary schools, teachers might have been uncertain about the function mapping their effort into students' ECE test scores. The connection between teachers' effort and their expected benefit might have therefore been diluted, making the incentive insufficient for prompting teachers into exerting more effort (Fryer, 2013).[55] Even

[54] In the study of Glewwe et al. (2010), for example, the survey to teachers was also framed as soliciting feedback on the incentive program; teachers in the treatment group were more likely to report having increased the number of homework assignments, whereas student reports suggest no such differences. In Behrman et al. (2015), teachers were also more likely to report that they spent more hours preparing their students for the test, although the incentive had no impact on student outcomes.

[55] A series of experimental studies in rural India suggest that teachers' knowledge and incentives might be complementary inputs in the education production function. While Muralidharan

if teachers knew how to equip their students with the skills needed to obtain high ECE scores, they might have encountered difficulties in passing on the incentive to their students. Since ECE tests have no impact whatsoever on students' grades, and the Ministry of Education only reports school averages (and not individual test scores) to schools, teachers, parents and even students, the latter might have little or no incentive to put effort in these tests.[56] Teachers might have anticipated that their actions would only marginally impact their students' ECE scores, and thus might have been discouraged from exerting more effort. The results from the experimental study implemented by Behrman et al. (2015) in Mexico provide suggestive evidence on the possibility that incentivizing teachers on their students' performance might not be effective unless students have a stake as well.[57] This hypothesis is partially supported by our online survey to teachers. When asked whether they thought students put any effort when taking the ECE test, 37% of survey respondents who taught math or language in 8^{th} grade answered that they did not. We inquired about the reasons for why students do not put any effort while taking the ECE, and teachers replied that this was due to the fact that ECE test scores do not affect their final grade (51%), because students are unmotivated (47%), the test is too long (10%) or too difficult (8%), and students are not familiar with these types of evaluations (11%).

Since the ECE was implemented in secondary schools for the first time in 2015, schools might not have known the relative standing of their students in

and Sundararaman (2010) show that giving teachers feedback on their students' past performance and detailed information on how to improve students' learning in low stakes tests has no effect on tests scores, students' test scores increased when this informational treatment was paired with a monetary incentive to teachers based on the performance of their students (the treatment from Muralidharan and Sundararaman, 2011). Since there is no treatment arm with monetary incentives but no information, it is hard to disentangle whether this effect is simply due to the monetary incentives, or whether the latter are only effective when teachers are given enough information on how to influence student learning.

[56] The findings of Tran and Zeckhauser (2012) and Azmat and Iriberri (2010) are consistent with the notion that not informing students about their achievement in the ECE might keep them from applying themselves while taking the test. Both studies find that providing students with relative performance feedback enhances their performance, even if they are not rewarded for it.

[57] The evidence provided by Behrman et al. (2015) on this point is only suggestive because, as compared to the treatment in which only teachers were incentivized, the potential reward for teachers and students was larger in the treatment arm in which both were incentivized. It is therefore hard to tease out if this incremental effect is coming the existence of complementarities between teachers' and students' effort, or from the fact that the monetary incentives were larger.

comparison to those from competing schools. Teachers might have been unable to infer the level of effort needed for their school to win the bonus, thus lowering the power of the incentive and ultimately discouraging them from putting in more effort. Since schools participating in the BE compete against other comparable schools within their district, they might have some prior about how their students compare to those of the competing schools, especially in BE groups with few schools. As shown in columns (5) and (10) of Table 1.9, there is a small but positive effect (0.012 and 0.023 SD in math and language) on student learning in schools in BE groups smaller than 27, the median group size. Although group size is probably an inaccurate proxy for knowledge about the probability of winning, this suggests that it might be important for schools to know how much effort they need to exert for the program to be effective.[58]

1.7.6 Teachers Did Not Have Enough Time to React

Finally, schools might not have had enough time to increase their students' learning in a meaningful way. As explained in Section 1.2, the Minister of Education mentioned the possibility of extending BE to secondary schools at the end of 2014, but the programs' regulation and the Ministry of Education's corresponding diffusion campaign only came out in July 2015, four months before the November 2015 ECE. In our survey to teachers, of those who taught 8^{th} grade and knew about BE in 2015, 39% reported that they heard about the program in the first trimester, 26% in the second, and 33% in the third (and 2% could not remember when they found out). The programs implemented by Muralidharan and Sundararaman (2011), Glewwe et al. (2010) and Lavy (2009) were announced 7-8 months before students were tested. Even though these papers find positive and sizable effects in this short time frame, teachers in Peru might have had less time to react, especially those who found out about the program later in the year. Fur-

[58] One of the questions in our survey shows a hypothetical ranking of 20 urban schools from the same school district, and asks teachers to identify what position would be held by a school with the same characteristics as the one they work for, and how that position would change if every teacher in their school dedicated an extra hour a day to improve the performance of their students (extra tutoring sessions, training sessions, etc.). In 47 % of cases, math and language teachers in 8^{th} grade answered that their school would still be below the 80^{th} percentile (i.e., would not win the bonus) after everyone changed their pedagogical practices.

thermore, when asked whether they thought there was enough time for students to improve their performance in the ECE in 2015 from the moment they found out about BE until the test, 67% of 8^{th} grade teachers who knew about the BE answered that there was not enough time.[59]

1.8 Conclusion

Can tying teachers' pay to the performance of their students improve their learning? We examine the impact of a collective teacher pay-for-performance program (*Bono Escuela*) implemented in 2015 in all public secondary schools in Peru, and find that it had no impact on students' math and language internal grades. Our coefficients are precisely estimated, allowing us to reject effects larger than 0.017 standard deviations, well below those previously found in the literature. Moreover, we find no evidence that the teacher incentive program had differential effects over schools or students of certain characteristics. We stipulate that the lack of increase in student learning might have been triggered by certain aspects of the evaluation linked to the bonus (students' low stakes and teachers' inexperience with it). These factors, along with schools' uncertainty about their potential ranking might have discouraged teachers from exerting higher effort. Finally, we argue that the program's timing might have played a role, possibly leaving teachers with insufficient time to instill significant learning gains in their students.

All in all, the results from our study suggest that successfully scaling up teacher pay-for-performance requires a deeper understanding about the role played by the different characteristics of these programs in their success. In particular, our findings raise the question of whether the interaction between between teachers' incentives and their information is important for these programs to work. If these complementarities exist, the efficacy of teacher incentives might depend on whether they are paired with teacher training. This paper also points to the fact that the type of exam being incentivized, and particularly the stakes that stu-

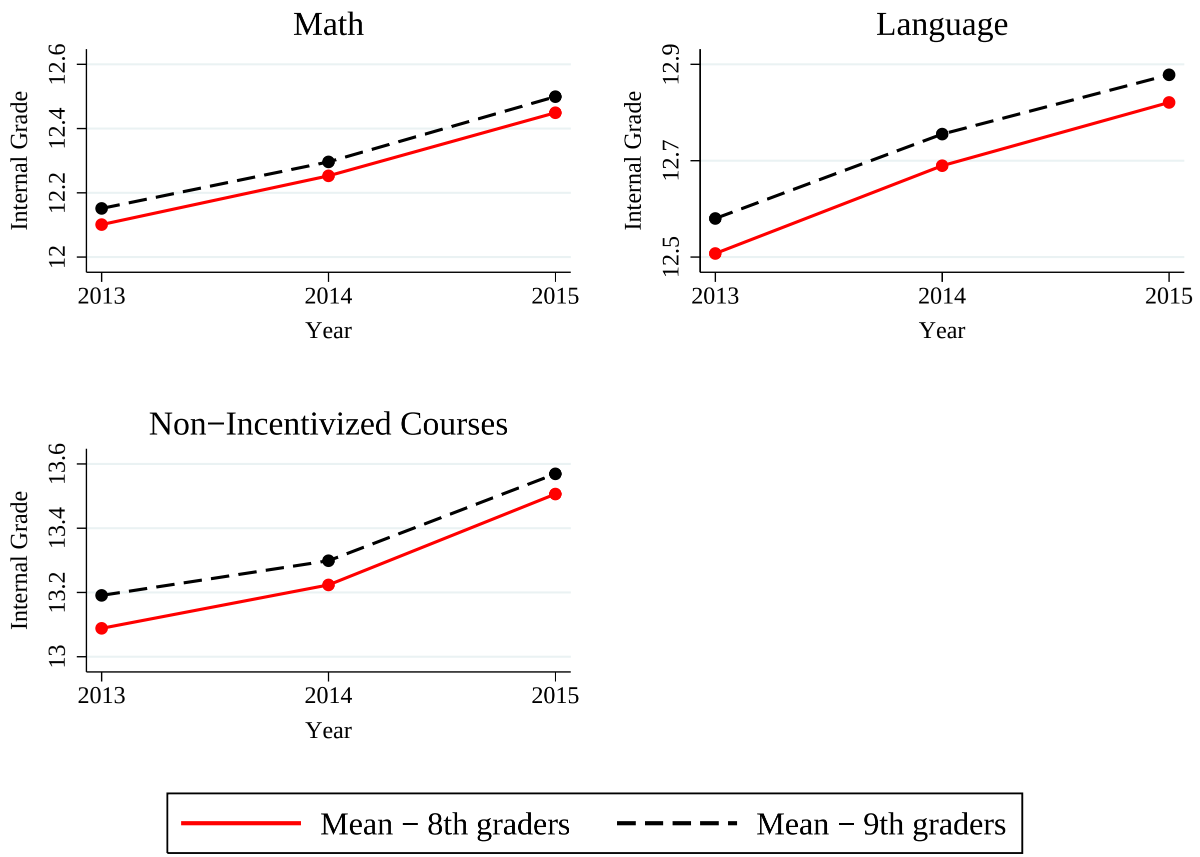
[59] While this could be an ex-post justification, we cannot discard that lack of time is one of the reasons the program had no impact. The future analysis of students' performance in the 2016 wave of the BE, for which teachers have the entire school year to prepare, might allow us to elucidate if this is potentially one of the reasons why the program had no effect on 8^{th} graders' performance in 2015.

dents have in it, might be important for teacher pay-for-performance programs to raise student learning. Going forward, research on teacher incentives should experimentally examine the complementarities between teachers' incentives, their knowledge, and their students' stakes in the incentivized outcome.

The fact that BE had no effect in the short-term does not imply that the program would have the same learning impacts if extended for a longer period. For one, teachers would acquire more experience with both the ECE and the BE. Consequently, some of the potential issues that could be diluting the effect of the incentive program may disappear. For instance, teachers would have more time to react to the incentive, would be more familiar with the test, and schools would have more information about their potential ranking within the group of schools they are competing against. Furthermore, teachers might only find it worthwhile to make sizable investments in improving their pedagogy if the program is continued and not only a one-off event. On the other hand, the program could have undesirable long-run impacts if teachers become more acquainted with how to teach-to-the-test, or if schools divert resources away from students not reached by the ECE. Extending the program could also result in schools devoting higher effort to improving the learning of 7^{th} grade students, in anticipation of their participation in the ECE standardized test in the following year. We plan to study these issues in our future research, once students' achievement data from 2016 becomes available. Finally, the program could affect the quality of teachers attracted to public schools, impacting the performance of students in the entire school. Although there is some evidence on the role of financial incentives in shaping the attributes of candidates for public sector jobs (e.g., Dal Bó et al., 2013 and Deserranno, 2016),[60] this question has not been tackled in the context of teacher pay-for-performance programs yet.

[60] While Dal Bó et al. (2013) find that higher wages for advertised government jobs in Mexico attract candidates with higher capabilities and greater motivation for working in the public sector, Deserranno (2016) finds that higher financial incentives for health promoters in Uganda attract more candidates, but hamper retention and performance because people drawn to the position are less likely to have pro-social preferences.

Figure 1.1: Trend in Average Internal Grades for 8^{th} and 9^{th} Graders



Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools which were eligible for taking the 2015 ECE standardized test and which are registered in the Ministry of Education's SIAGIE administrative system. The figures plot the average of all 8^{th} and 9^{th} graders internal grades in math, language and non-incentivized courses, respectively. We take the average of non-incentivized courses, which are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce.

Figure 1.2: Timing of BE in Secondary Schools

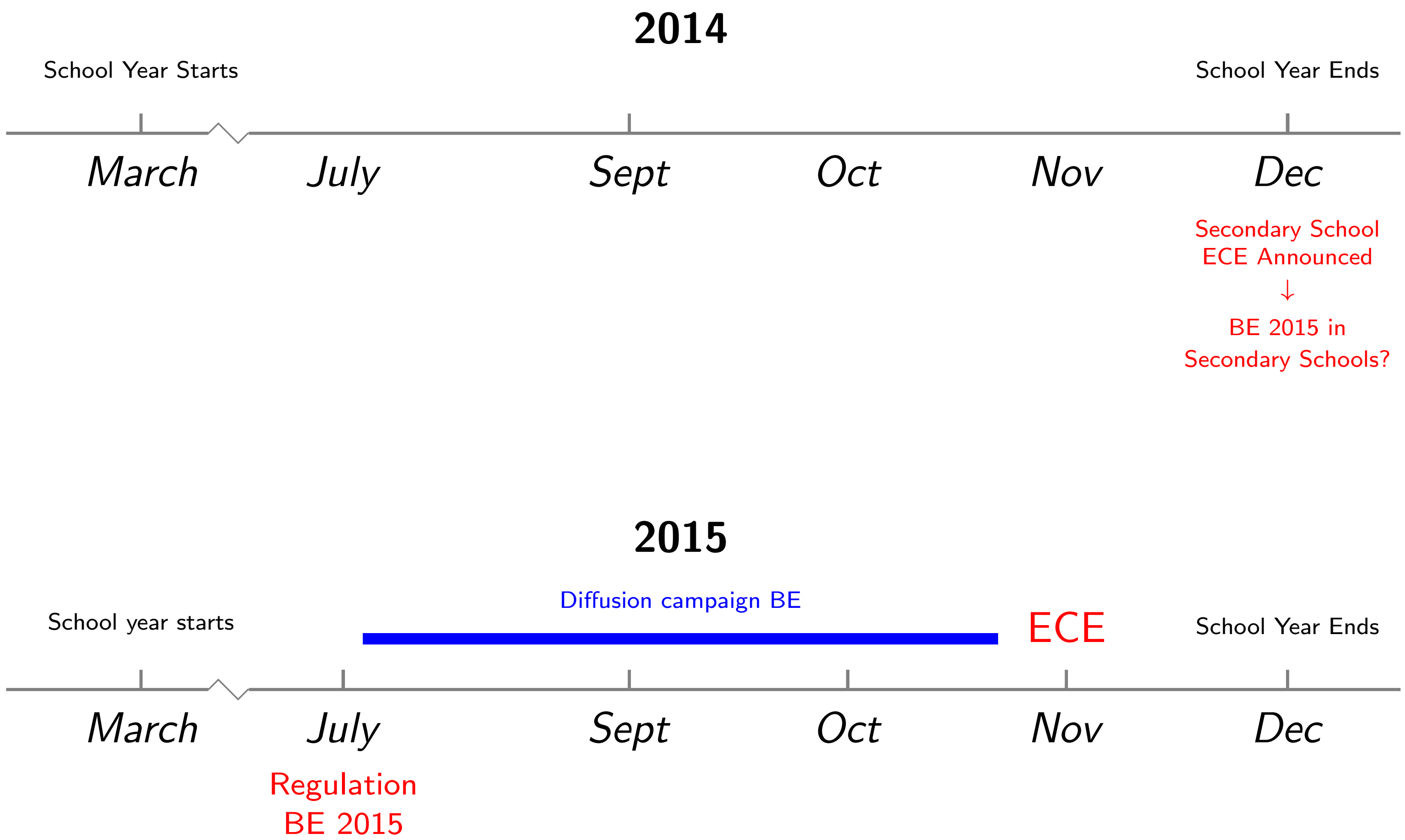
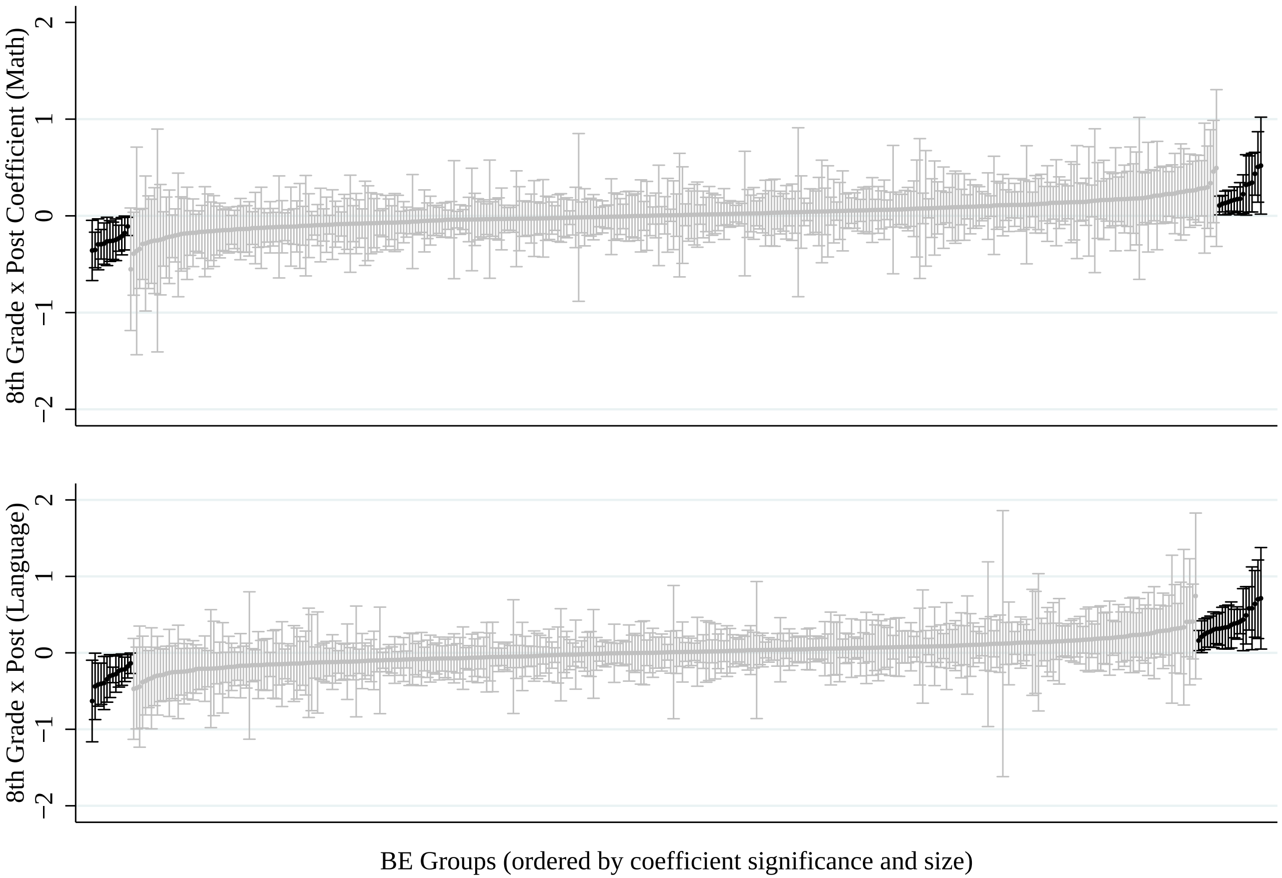
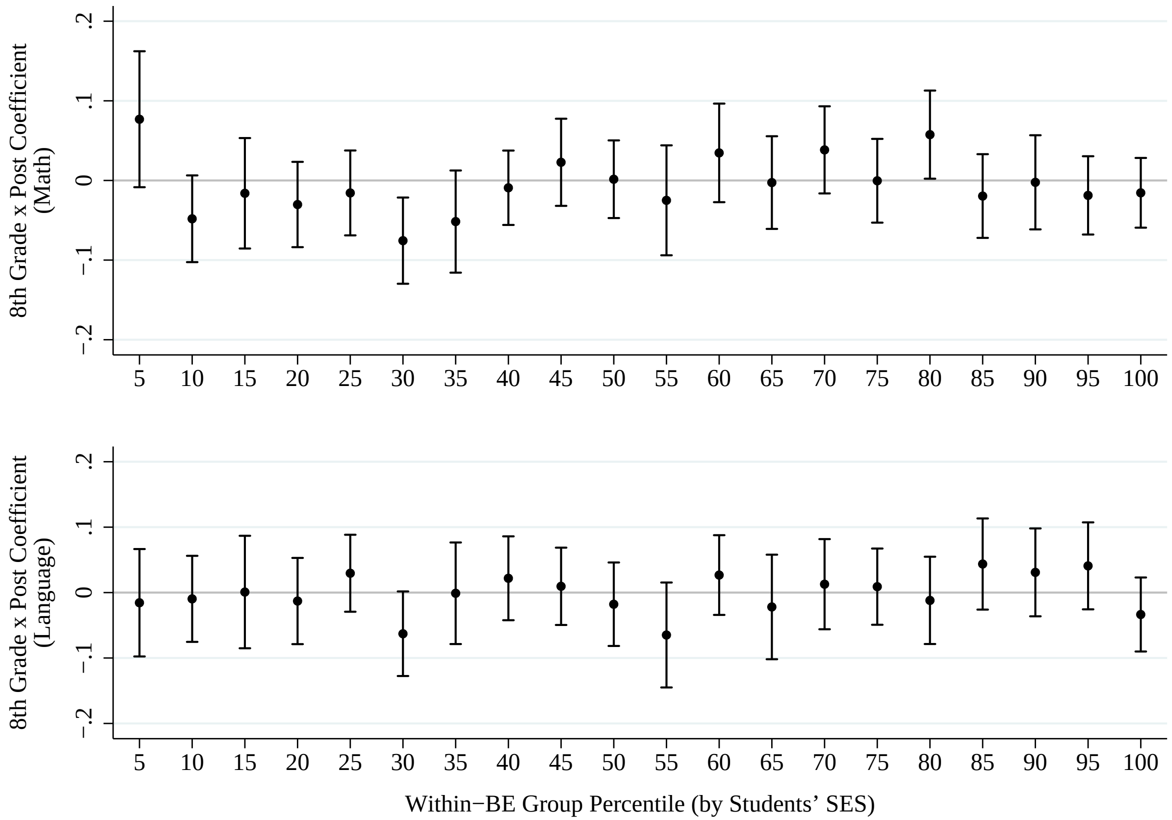


Figure 1.3: Effect of Teacher Incentive on Students' Math and Language Internal Grades in each BE Group



Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools which were eligible for taking the 2015 ECE standardized test and which are registered in the Ministry of Education's SIAGIE administrative system. The figures plot the 8^{th} *Grade* x *Post* coefficients and their 95% confidence intervals separately estimated for each BE group in math and language, respectively. BE groups in both figures are ordered by significance and coefficient size, and the ordering is separately done in each figure.

Figure 1.4: Heterogeneous Effect of Teacher Incentive on Students' Internal Grades by Schools' SES Rank



Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variables in the top and bottom figures are students' internal grades in math and language, respectively, standardized by course-year. We construct a SES index (taking values 0-3) for each 8^{th} student in 2015, adding up three dummies for whether his first language is Spanish, and whether his mother and father have more than a primary school education. We then calculate the average index for each school, and group schools in 20 percentile groups according to their ranking in terms of this measure within their BE group. The figures plot the coeffient and 95% confidence interval for the interaction of each percentile dummy and the 8^{th} x Post dummy from our baseline regressions.

Table 1.1: Assignment of Score in BE

Weight	Indicator	Relevant Grades
40%	Average math and language score in 2015 ECE standardized tests	8th Grade
35%	Intra-annual retention rates	All Grades
5%	Enrollment of students in SIAGIE administrative system	All Grades
12%	Teacher attendance, management of school infrastructure and compliance with class hours	All Grades
8%	Pedagogical practices and learning environment	8th Grade

Source: Decree 203-2015.

Table 1.2: Summary Statistics for 8^{th} and 9^{th} Graders

	8^{th} Grade		9^{th} Grade	
	Mean	Std. Dev	Mean	Std. Dev
Final Grade (0-20)				
Math	12.27	2.17	12.32	2.17
Language	12.67	2.07	12.74	2.07
Other courses - average	13.27	1.60	13.35	1.60
Passed the Course				
Math	0.84	0.37	0.85	0.36
Language	0.89	0.31	0.90	0.30
Other courses - average	0.93	0.15	0.94	0.14
Other Individual Characteristics				
Male	0.49	0.50	0.50	0.50
Repeated last year	0.06	0.23	0.04	0.20
Foreigner	0.00	0.05	0.00	0.05
Spanish is native tongue	0.84	0.37	0.83	0.38
Has a disability	0.00	0.06	0.00	0.06
Father is alive	0.90	0.30	0.89	0.31
Mother is alive	0.97	0.16	0.97	0.17
Father lives in HH	0.76	0.43	0.77	0.42
Mother lives in HH	0.80	0.40	0.80	0.40
Number of students	1,090,496		1,018,310	
Grade/School Characteristics				
Rural	0.41	0.49	0.40	0.49
Number of classes	2.00	1.92	1.94	1.84
Teacher-pupil ratio	19.61	8.72	18.98	8.75
Number of teachers	10.74	6.55	10.95	6.71
% of teachers instructing the other grade	0.83	0.22	0.83	0.20
Average age of teachers	41.64	5.34	41.66	5.26
% of male teachers	0.60	0.21	0.60	0.20
Number of school-year observations	23,810		23,469	

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. We exclude students for which we have no grades and/or no individual controls (0.4%). Since teacher data is missing for a small subsample of schools, the number of grade-observations for teacher characteristics is 23,462 and 23,127 in 8^{th} and 9^{th} grade. *Final Grade* is the students' internal grades at the end of the school year in math, language and non-incentivized courses. *Passed the Course* is a dummy for whether the student got an 11 or higher in that particular course (the requirement for passing). We take the average of non-incentivized courses, which are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce. *Repeated last year* is a dummy for whether the student was retained in the same grade at the end of the previous year. *Rural* is a dummy for whether the school is in a rural area, and *Number of classes* is the number of classes in the student's grade and year. *Number of teachers* is the total number of teachers in that grade and year, and *% of teachers instructing the other grade* is the % of 8^{th} (9^{th}) teachers also teaching 9^{th} (8^{th}) grade in the same school.

Table 1.3: Effect of Teacher Incentive on Students' Math and Language Internal Grades

	Math			Language		
	(1)	(2)	(3)	(4)	(5)	(6)
8^{th} Grade x Post	-0.001	-0.003	-0.005	0.006	0.004	0.001
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
8^{th} Grade	-0.022***	-0.016***	-0.008**	-0.034***	-0.026***	-0.016***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Repeated last year			-0.570***			-0.623***
			(0.006)			(0.007)
Male			0.115***			0.334***
			(0.003)			(0.003)
Foreigner			0.046**			0.089***
			(0.019)			(0.019)
Spanish is native tongue			0.171***			0.174***
			(0.007)			(0.007)
Has a disability			-0.262***			-0.256***
			(0.014)			(0.015)
Father is alive			0.066***			0.066***
			(0.004)			(0.003)
Mother is alive			0.040***			0.049***
			(0.006)			(0.006)
Father lives in HH			0.026***			0.020***
			(0.002)			(0.002)
Mother lives in HH			0.014***			0.019***
			(0.003)			(0.002)
Observations	2108806	2108806	2108806	2108793	2108793	2108793
R^2	0.000	0.071	0.092	0.000	0.087	0.135
Year FE	X	X	X	X	X	X
School FE		X	X		X	X
Individual Controls			X			X

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variables are students' internal grades in math and language, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.4: Effect of Teacher Incentive on Students' Internal Grades in Non-Incentivized Courses

	Non-Incentivized Courses		
	(1)	(2)	(3)
8^{th} Grade x Post	0.014***	0.014***	0.011***
	(0.004)	(0.004)	(0.004)
8^{th} Grade	-0.043***	-0.038***	-0.028***
	(0.003)	(0.003)	(0.002)
Repeated last year			-0.601***
			(0.006)
Male			0.284***
			(0.003)
Foreigner			0.047***
			(0.015)
Spanish is native tongue			0.120***
			(0.005)
Has a disability			-0.232***
			(0.012)
Father is alive			0.063***
			(0.003)
Mother is alive			0.044***
			(0.005)
Father lives in HH			0.022***
			(0.002)
Mother lives in HH			0.013***
			(0.002)
Observations	2108972	2108972	2108972
R^2	0.001	0.120	0.185
Year FE	X	X	X
School FE		X	X
Individual Controls			X

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variable is students' internal grades in non-incentivized courses, standardized by course-year. We first standardize each of the non-incentivized courses by course-year, and then take the average. Non-incentivized courses are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.5: Heterogeneous Effect of Teacher Incentive on Students' Internal Grades

	Male	Spanish Speaker	Parents High Educ	Repeated	Lagged Grade	Ln Enrollment	Num. Classes	Rural
Panel A: Math Grades								
8^{th} Grade x Post	-0.010	-0.004	-0.005	-0.003	-0.002	0.033	-0.002	-0.007
	(0.008)	(0.012)	(0.007)	(0.007)	(0.008)	(0.023)	(0.009)	(0.008)
8^{th} Grade x Post x Covariate	0.010	-0.000	-0.003	-0.019	0.006	-0.008	-0.001	0.016
	(0.009)	(0.013)	(0.005)	(0.016)	(0.005)	(0.006)	(0.002)	(0.012)
Observations	2108806	2108806	1851727	2108806	1382813	2108806	2108806	2108806
R^2	0.092	0.092	0.099	0.092	0.440	0.092	0.092	0.092
Panel B: Language Grades								
8^{th} Grade x Post	0.006	-0.000	-0.004	0.002	0.005	-0.023	-0.008	0.004
	(0.009)	(0.013)	(0.009)	(0.008)	(0.009)	(0.027)	(0.010)	(0.009)
8^{th} Grade x Post x Covariate	-0.010	0.002	0.003	-0.008	0.001	0.006	0.002	-0.017
	(0.010)	(0.015)	(0.006)	(0.016)	(0.005)	(0.007)	(0.002)	(0.014)
Observations	2108793	2108793	1851715	2108793	1382756	2108793	2108793	2108793
R^2	0.135	0.135	0.142	0.135	0.417	0.135	0.135	0.135

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. Heterogeneities by *Lagged Grade* exclude the year 2013 for which students' previous grade is unavailable, and heterogeneities by parents' education exclude 12% of students in 2013-2015 for which this variable is missing. The dependent variables are students' internal grades in math and language, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, *Post* is a dummy for the year 2015, and *Covariate* is the variable indicated in the column header. All regressions include school and year fixed effects, as well as the standard individual controls and the three-way interaction between 8^{th} *Grade*, *Post* and *Covariate*. We only report two coefficients for exposition purposes. *Spanish Speaker* is a dummy for whether the student's first language is Spanish, and *Parents High Educ* is 0 if both parents have a primary school degree or less, is 1 if only one of the parents has more than a primary school degree, and 2 if both. *Repeated* is a dummy for whether the student was retained in the same grade at the end of the previous year, and *Lagged Grade* is the students' internal grade in that particular course in the previous year, standardized by school and grade. *Ln Enrollment* is the log of the number of students enrolled in that year and grade. *Num. Classes* is the number of classes in the student's grade and year, *Rural* is a dummy for whether the school is in a rural area. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.6: Test for Parallel Trends in Students' Internal Grades

	Math	Language	Non-Incentivized Courses
8^{th} Grade x Post	-0.004	0.002	0.016***
	(0.007)	(0.009)	(0.005)
8^{th} Grade x 2014	0.002	0.001	0.010**
	(0.007)	(0.009)	(0.004)
8^{th} Grade	-0.009*	-0.017***	-0.033***
	(0.005)	(0.007)	(0.003)
Observations	2108806	2108793	2108972
R^2	0.092	0.135	0.185

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE are registered in the Ministry of Education's SIAGIE administrative system. All regressions include year fixed effects, school fixed effects, and the standard controls. The dependent variables are students' internal grades in math, language and non-incentivized courses, standardized by course-year. We take the average of non-incentivized courses, which are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, *Post* is a dummy for the year 2015, and *2014* is a dummy for the year 2014. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.7: Heterogeneity by Overlap of 8^{th} and 9^{th} Grade Teachers

	Math			Language		
	Low	Med	High	Low	Med	High
8^{th} Grade x Post	-0.014	-0.014	0.013*	0.010	0.001	-0.009
	(0.014)	(0.012)	(0.007)	(0.017)	(0.014)	(0.008)
8^{th} Grade	-0.001	-0.006	-0.016***	-0.017	-0.017**	-0.014***
	(0.008)	(0.007)	(0.004)	(0.011)	(0.008)	(0.005)
Repeated last year	-0.594***	-0.533***	-0.582***	-0.637***	-0.587***	-0.642***
	(0.013)	(0.008)	(0.009)	(0.014)	(0.009)	(0.010)
Male	0.123***	0.128***	0.097***	0.369***	0.367***	0.275***
	(0.006)	(0.005)	(0.004)	(0.007)	(0.005)	(0.005)
Foreigner	0.048*	0.017	0.124**	0.084***	0.082***	0.142**
	(0.027)	(0.028)	(0.055)	(0.027)	(0.029)	(0.057)
Spanish is native tongue	0.130***	0.180***	0.198***	0.129***	0.185***	0.202***
	(0.013)	(0.012)	(0.009)	(0.013)	(0.013)	(0.009)
Has a disability	-0.231***	-0.247***	-0.300***	-0.200***	-0.243***	-0.310***
	(0.028)	(0.025)	(0.019)	(0.026)	(0.027)	(0.021)
Father is alive	0.070***	0.068***	0.061***	0.073***	0.061***	0.065***
	(0.007)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)
Mother is alive	0.020	0.044***	0.055***	0.029**	0.051***	0.067***
	(0.013)	(0.010)	(0.008)	(0.011)	(0.010)	(0.009)
Father lives in HH	0.034***	0.032***	0.011***	0.028***	0.023***	0.006
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Mother lives in HH	0.013***	0.012***	0.020***	0.015***	0.017***	0.025***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	740453	672268	696085	740446	672275	696072
R^2	0.072	0.084	0.126	0.116	0.129	0.166
Avg. % of teachers in both grades	0.392	0.654	0.906	0.392	0.654	0.906

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. Columns Low, Med and High restrict the sample to students in schools with a low, medium and high average overlap between 8^{th} and 9^{th} grade teachers in 2013-2015. Overlap between 8^{th} and 9^{th} grade teachers is the % of teachers in 8^{th} grade also instructing in 9^{th} grade. The dependent variables are students' internal grades in math and language, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.8: Test for Changes in Teacher Composition Across Grades

	Average Age	% Male	Average Number of Classes	Average Number of Schools	% New to School	% New to School-Grade
8^{th} Grade x Post	-0.013	0.001	0.016	-0.005**	-0.000	-0.000
	(0.021)	(0.001)	(0.016)	(0.002)	(0.001)	(0.002)
8^{th} Grade	0.043***	-0.006***	-0.090***	0.000	0.001	0.001
	(0.013)	(0.001)	(0.011)	(0.001)	(0.001)	(0.001)
Observations	46614	46615	46615	46615	31332	31332
R^2	0.869	0.738	0.953	0.737	0.808	0.758
Mean Dep. Variable	41.685	0.598	11.790	1.634	0.474	0.552

Notes: The sample includes all public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE standardized test, registered in the Ministry of Education's SIAGIE administrative system, and with data on teacher characteristics. The unit of analysis in these regressions is a school-grade-year, for 8^{th} and 9^{th} grade. *Average Age* is the average age of teachers in that grade, and *% Male* is the % of teachers in that grade that are male. *Average Number of Classes* is the average number of courses taught by teachers in that grade, and *Average Number of Schools* is the mean number of different secondary schools in which the teacher works. *% New to School-Grade* are the proportion of teachers in that particular grade who are new to the school, or new to that particular grade , respectively. All regressions include school fixed effects, year fixed effects, a dummy for 8^{th} *Grade*, and the interaction between 8^{th} *Grade* and a dummy for 2015 (i.e., *Post*). The regressions in columns 5 and 6 do not include the year 2013 since we do not have information on teachers' appointments in 2012. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.9: Effect of Teacher Incentive by Average Salary, Number of Classes, School's Experience with Primary School BE and BE Group Size

	Math					Language				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
8^{th} Grade x Post	0.249	-0.011	-0.006	-0.007	-0.014*	0.299	0.000	0.005	-0.000	-0.010
	(0.501)	(0.008)	(0.009)	(0.007)	(0.008)	(0.585)	(0.010)	(0.011)	(0.009)	(0.011)
8^{th} Grade x Post x Ln (Average Salary)	-0.035					-0.041				
	(0.069)					(0.081)				
8^{th} Grade x Post x One Class		0.030***					0.004			
		(0.011)					(0.013)			
8^{th} Grade x Post x BE Primary			0.003					-0.007		
			(0.013)					(0.016)		
8^{th} Grade x Post x BE Primary Winner				0.016					0.010	
				(0.018)					(0.024)	
8^{th} Grade x Post x Small BE Group					0.026**					0.033**
					(0.013)					(0.016)
Observations	2077227	2108806	2108806	2108806	2108806	2077214	2108793	2108793	2108793	2108793
R^2	0.092	0.092	0.092	0.092	0.092	0.135	0.135	0.135	0.135	0.135
P-Value (sum of both coefficients)	0.621	0.008	0.739	0.568	0.237	0.601	0.558	0.893	0.669	0.062

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variables are students' internal grades in math and language, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, *Post* is a dummy for the year 2015, *Ln (Average Salary)* is the average salary of teachers in each school in 2015 (in logs), obtained from the 2015 school census, and *One Class* is a dummy for whether the student attends a school in which there is only one class in his grade. *BE Primary* is a dummy for whether a primary school that participated in the BE in the past operates in the same building, and *BE Primary* is a dummy for whether there is a primary school in the building that won the BE bonus in the past. *Small BE Group* is a dummy for whether the number of schools in the corresponding BE group is below the median. All regressions include year and school fixed effects, the standard individual controls, and the three-way interaction between 8^{th} *Grade*, *Post* and the specific heterogeneity variable. We only report two coefficients for exposition purposes. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.10: Effect of Teacher Incentive on Teachers' Pedagogical Practices

	8^{th} Grade	Other Grades	Difference	P-Value
Panel A: All Math/Language Teachers				
Improved attendance	0.207	0.157	0.050**	0.024
More homework, evaluations and/or tutoring sessions	0.471	0.370	0.101***	0.000
Paid more attention to weakest students	0.683	0.637	0.046*	0.097
Training programs or feedback sessions	0.548	0.542	0.006	0.828
Increased difficulty of classes	0.192	0.135	0.056***	0.007
Decreased difficulty of classes	0.148	0.138	0.010	0.620
More multiple choice tests	0.385	0.299	0.086***	0.002
Other	0.130	0.150	-0.020	0.317
Number of teachers	454	865		
Panel B: Math/Language Teachers in Both Grades				
Improved attendance	0.203	0.143	0.060***	0.000
More homework, evaluations and/or tutoring sessions	0.460	0.326	0.134***	0.000
Paid more attention to weakest students	0.677	0.657	0.020	0.209
Training programs or feedback sessions	0.523	0.494	0.029	0.149
Increased difficulty of classes	0.197	0.157	0.040**	0.016
Decreased difficulty of classes	0.146	0.131	0.014	0.298
More multiple choice tests	0.391	0.337	0.054**	0.017
Other	0.123	0.143	-0.020	0.250
Number of teachers	350	350		

Notes: The sample includes all survey respondents who taught math or language in 8^{th} and other grades in 2015, and knew about the BE program during the 2015 academic year. Panel B only includes those who taught math or language in 8^{th} grade and other grades. Teachers were asked whether they changed their pedagogical practices in 2015 as a result of BE, and could answer any of the options specified in the table rows. We asked them separately about changes while teaching 8^{th} grade (column 1) as opposed to all other grades (column 2), in case the teacher taught both. * significant at 10%; ** significant at 5%; *** significant at 1%

1.9 Appendix Figures and Tables

Table 1.A1: Test for Parallel Trends Comparing Public and Private Schools

	Math	Language	Non-Incentivized Courses
Public x Post	0.116***	0.100***	0.143***
	(0.010)	(0.011)	(0.007)
Public x 2014	0.044***	0.063***	0.036***
	(0.008)	(0.010)	(0.006)
Repeated last year	-0.527***	-0.586***	-0.548***
	(0.006)	(0.006)	(0.005)
Male	0.109***	0.323***	0.271***
	(0.003)	(0.003)	(0.002)
Foreigner	0.037***	0.058***	0.061***
	(0.012)	(0.011)	(0.009)
Spanish is native tongue	0.183***	0.193***	0.126***
	(0.007)	(0.007)	(0.005)
Has a disability	-0.247***	-0.265***	-0.232***
	(0.014)	(0.015)	(0.013)
Father is alive	0.073***	0.075***	0.068***
	(0.003)	(0.004)	(0.003)
Mother is alive	0.031***	0.038***	0.039***
	(0.006)	(0.006)	(0.005)
Father lives in HH	0.037***	0.031***	0.030***
	(0.002)	(0.002)	(0.002)
Mother lives in HH	0.011***	0.016***	0.011***
	(0.002)	(0.002)	(0.002)
Observations	1514619	1514593	1514717
R^2	0.155	0.196	0.293

Notes: The sample includes all 8^{th} grade students in 2013-2015, in public and private schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. All regressions include year fixed effects, school fixed effects, and the standard controls. The dependent variables are students' internal grades in math, language and non-incentivized courses, standardized by course-year. We take the average of non-incentivized courses, which are art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce. *Public* is a dummy for whether the students attends a public school, *Post* is a dummy for the year 2015, and *2014* is a dummy for the year 2014. The coefficient for the *Public* dummy is not display since this variable is perfectly collinear with the corresponding school fixed effect. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.A2: Effect of Teacher Incentive on Students' Grades in Non-Incentivized Courses

	Arts	Science	Social Studies	English	Civics	Human Relations	Physical Education	Religion	Educ. for the Workforce
8^{th} Grade x Post	0.003	-0.001	0.017*	0.011	0.008	0.019**	0.012	0.014*	0.012
	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
8^{th} Grade	-0.037***	0.052***	-0.039***	-0.046***	-0.055***	-0.044***	-0.016***	-0.042***	-0.027***
	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Repeated last year	-0.621***	-0.570***	-0.602***	-0.573***	-0.598***	-0.588***	-0.617***	-0.611***	-0.624***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)	(0.008)
Male	0.336***	0.236***	0.259***	0.281***	0.339***	0.384***	0.074***	0.393***	0.256***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
Foreigner	0.029*	0.057***	0.039**	0.163***	0.041**	0.030	0.061***	-0.015	0.003
	(0.017)	(0.018)	(0.020)	(0.019)	(0.020)	(0.018)	(0.015)	(0.018)	(0.018)
Spanish is native tongue	0.088***	0.148***	0.128***	0.147***	0.128***	0.138***	0.092***	0.103***	0.107***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)
Has a disability	-0.188***	-0.261***	-0.229***	-0.276***	-0.235***	-0.239***	-0.257***	-0.185***	-0.219***
	(0.016)	(0.014)	(0.016)	(0.014)	(0.015)	(0.014)	(0.015)	(0.016)	(0.016)
Father is alive	0.064***	0.063***	0.065***	0.068***	0.063***	0.066***	0.052***	0.062***	0.064***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Mother is alive	0.050***	0.043***	0.040***	0.050***	0.038***	0.043***	0.047***	0.039***	0.043***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Father lives in HH	0.024***	0.026***	0.024***	0.020***	0.025***	0.023***	0.013***	0.022***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mother lives in HH	0.012***	0.014***	0.013***	0.015***	0.014***	0.014***	0.012***	0.016***	0.010***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	2108795	2108792	2108780	2108791	2108793	2108777	2108794	2071576	2108778
R^2	0.196	0.127	0.138	0.143	0.154	0.168	0.235	0.183	0.186

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2013-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. The dependent variables are students' internal grades in art, science, social studies, English, civics, human relations, physical education, religion, and education for the workforce, standardized by course-year. 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.A3: Non-Linear Heterogeneous Effects by Students' Lagged Grade

	Math	Language
Panel A: Lagged Grade Quartiles		
8^{th} Grade x Post	-0.003	0.003
	(0.009)	(0.011)
8^{th} Grade x Post x Q2	-0.006	0.009
	(0.009)	(0.010)
8^{th} Grade x Post x Q3	-0.010	0.008
	(0.010)	(0.011)
8^{th} Grade x Post x Q4	0.014	0.001
	(0.014)	(0.015)
Observations	1382813	1382756
R^2	0.393	0.387
P-value (sum of coefficients Q2)	0.346	0.281
P-value (sum of coefficients Q3)	0.198	0.349
P-value (sum of coefficients Q4)	0.380	0.792
Panel B: Lagged Grade Terciles		
8^{th} Grade x Post	-0.004	-0.004
	(0.009)	(0.010)
8^{th} Grade x Post x T2	-0.014	0.010
	(0.009)	(0.009)
8^{th} Grade x Post x T3	0.014	0.011
	(0.013)	(0.012)
Observations	1382813	1382756
R^2	0.359	0.363
P-value (sum of coefficients T2)	0.050	0.622
P-value (sum of coefficients T3)	0.444	0.588

Notes: The sample includes all 8^{th} and 9^{th} grade students attending public secondary school in 2014-2015, in public schools eligible for taking the 2015 ECE and registered in the Ministry of Education's SIAGIE administrative system. We exclude the year 2013 for which students' previous grade is unavailable. The dependent variables are students' internal grades in math and language, standardized by course-year. Students in Panel A (B) are divided into quartiles (terciles) according to their lagged grade (i.e., their internal grade in that particular course in the previous year, standardized by school and grade). 8^{th} *Grade* is a dummy for whether the students is in 8^{th} grade, and *Post* is a dummy for the year 2015. All regressions include school and year fixed effects, as well as the standard individual controls and the three-way interaction between 8^{th} *Grade*, *Post* and the Quartile or Tercile dummies. We only report the triple interactions for exposition purposes. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.A4: Cross-Sectional Correlation Between Average ECE Test Scores and Internal Grades in 2015

	Math ECE		Language ECE	
	Average (z-score)	% Very Low	Average (z-score)	% Very Low
Average internal grade (z-score)	0.116***		0.103***	
	(0.010)		(0.008)	
Failed course (% of students)		0.071***		0.091***
		(0.018)		(0.021)
Spanish as native tongue (% of students)	0.667***	-0.180***	0.836***	-0.250***
	(0.044)	(0.012)	(0.039)	(0.012)
Mother with high education (% of students)	0.930***	-0.223***	1.143***	-0.221***
	(0.078)	(0.020)	(0.065)	(0.018)
Father with high education (% of students)	0.438***	-0.120***	0.558***	-0.175***
	(0.075)	(0.020)	(0.060)	(0.018)
Repeated last year (% of students)	-0.138	0.044	0.079	-0.053
	(0.156)	(0.038)	(0.118)	(0.039)
Male (% of students)	-0.171***	0.037***	0.086*	-0.029**
	(0.054)	(0.014)	(0.046)	(0.013)
Teacher-pupil ratio	0.010***	-0.002***	0.010***	-0.002***
	(0.001)	(0.000)	(0.001)	(0.000)
Long school-day	0.309***	-0.074***	0.187***	-0.052***
	(0.024)	(0.006)	(0.018)	(0.005)
Rural	-0.050**	0.019***	-0.135***	0.049***
	(0.023)	(0.006)	(0.018)	(0.006)
Observations	8010	8010	8010	8010
R^2	0.501	0.491	0.684	0.617

Notes: The sample includes all public secondary schools taking the ECE in 2015 and registered in the Ministry of Education's SIAGIE administrative system, and the unit of observation is a school in the year 2015. The dependent variable in columns 1 and 3 are 8^{th} graders' average ECE grades in math and language, expressed as a z-score. The dependent variable in columns 2 and 4 is the % of students with very low achievement in the 2015 ECE. *Average internal grade (z-score)* is the school's average internal grade for 8^{th} grade students in 2015 , standardized across schools, in math (column 1) and language (column 3). *Failed course* measures the % of 8^{th} graders that failed math (column 2) and language (column 4) in 2015 according to their internal grades. *Mother with high education* and *Father with high education* indicate the % of 8^{th} graders in that school whose mother/father had more than a primary school degree in 2015. *Teacher-pupil-ratio* is the average number of 8^{th} grade students per class in 2015, and *Long school-day* is a dummy for whether the school had a longer instruction day in 2015. All regressions include school district fixed effects. Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.A5: Cross-Sectional Correlation Between School Covariates and Average Learning Outcomes in 2015

	Math		Language	
	ECE	Internal	ECE	Internal
Spanish as native tongue (% of students)	0.672***	0.044	0.857***	0.206***
	(0.045)	(0.056)	(0.039)	(0.054)
Mother with high education (% of students)	0.959***	0.252***	1.180***	0.363***
	(0.079)	(0.089)	(0.066)	(0.092)
Father with high education (% of students)	0.474***	0.314***	0.591***	0.327***
	(0.076)	(0.084)	(0.061)	(0.083)
Repeated last year (% of students)	-0.438***	-2.578***	-0.206*	-2.772***
	(0.157)	(0.168)	(0.117)	(0.170)
Male (% of students)	-0.132**	0.334***	0.125***	0.385***
	(0.055)	(0.065)	(0.046)	(0.065)
Teacher-pupil ratio	0.008***	-0.023***	0.008***	-0.020***
	(0.001)	(0.002)	(0.001)	(0.002)
Long school-day	0.338***	0.242***	0.214***	0.269***
	(0.024)	(0.030)	(0.018)	(0.030)
Rural	-0.048**	0.023	-0.135***	-0.001
	(0.023)	(0.028)	(0.018)	(0.028)
Observations	8010	8010	8010	8010
R^2	0.491	0.243	0.676	0.239

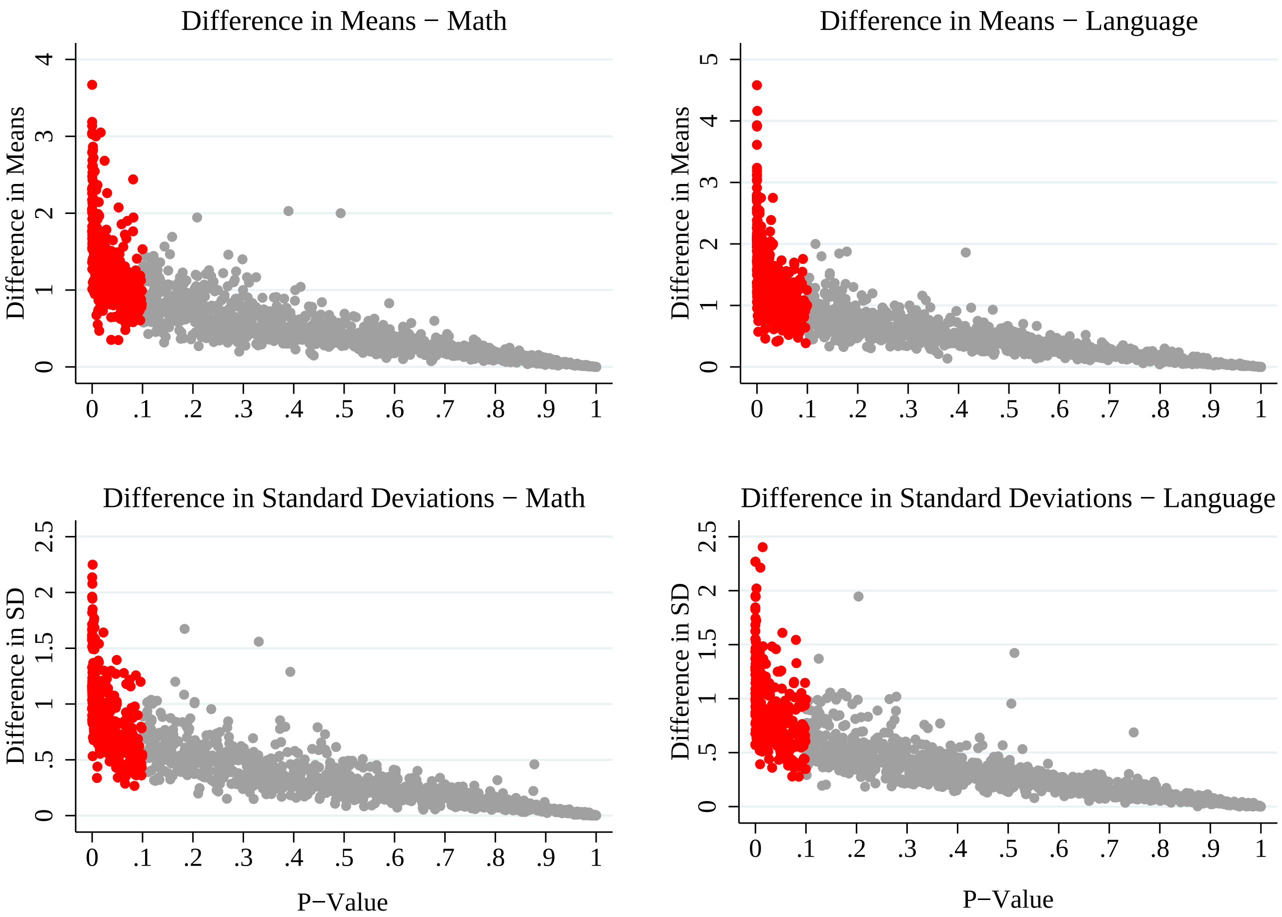
Notes: The sample includes all public secondary schools taking the ECE in 2015 and registered in the Ministry of Education's SIAGIE administrative system, and the unit of observation is a school in the year 2015. The dependent variable in columns 1 and 3 are 8^{th} graders' average ECE grades in math and language, whereas the dependent variable in columns 2 and 4 are the school's average internal grade for 8^{th} grade students in 2015. Average ECE and internal grades are standardized across schools (i.e., expressed as a z-score). *Mother with high education* and *Father with high education* indicate the % of 8^{th} graders in that school whose mother/father had more than a primary school degree in 2015. *Teacher-pupil-ratio* is the average number of 8^{th} grade students per class in 2015, and *Long school-day* is a dummy for whether the school had a longer instruction day in 2015. All regressions include school district fixed effects. Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.A6: Within-School Correlation Between Covariates and Learning Outcomes in 2015

	Math		Language	
	Grade (z-score)	Low Achievement	Grade (z-score)	Low Achievement
Panel A: ECE Grades				
Socioeconomic status index	0.135***	-0.044***	0.190***	-0.049***
	(0.003)	(0.001)	(0.003)	(0.001)
Male	0.220***	-0.071***	0.020***	-0.010***
	(0.004)	(0.002)	(0.004)	(0.002)
Observations	354429	354547	354529	354547
R^2	0.020	0.216	0.015	0.254
Panel B: Internal Grades				
Spanish as native tongue	0.204***	-0.016***	0.218***	-0.011***
	(0.012)	(0.004)	(0.012)	(0.003)
Mother with high education	0.158***	-0.021***	0.166***	-0.014***
	(0.005)	(0.002)	(0.005)	(0.001)
Father with high education	0.129***	-0.017***	0.136***	-0.014***
	(0.005)	(0.002)	(0.005)	(0.001)
Male	0.138***	-0.035***	0.353***	-0.051***
	(0.005)	(0.002)	(0.005)	(0.001)
Observations	324696	325320	324689	325320
R^2	0.019	0.103	0.044	0.099

Notes: Panel A contains the anonymized sample of 8^{th} graders taking the ECE standardized test in 2015, and the sample from Panel B includes all 8^{th} graders in 2015 from public secondary schools taking the ECE in 2015 and registered in the Ministry of Education's SIAGIE administrative system. The dependent variable in columns 1 and 3 of Panel A (Panel B) is the students' ECE (internal) grade in math and language, standardized by subject and school. The dependent variable in columns 2 and 4 of Panel A (Panel B) is a dummy for whether the student scored in the lowest category in the ECE (failed according to his internal grades). *Socioeconomic status* is an index ranging between -3.5 and 9.5, which is increasing in measures of socioeconomic status such as parents' education, and household assets and characteristics. All regressions include school fixed effects. Standard errors clustered by school in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1.A1: Variation in Internal Grades Across 8^{th} Grade Classes in 2014



Notes: Top figures depict the difference in means and the corresponding p-value when testing whether the mean math and language internal grades differ across 8^{th} grade classes from the same school in 2014, in every school with only two 8^{th} grade classes. The bottom figures depict the difference in the standard deviation and the corresponding p-value for an F-test of difference in variances in the same sample of schools.

Hoffman, Mitchell, Gianmarco León, and María Lombardi, "Compulsory Voting, Turnout and Government Spending: Evidence from Austria", *Journal of Public Economics*, 2017, 145, 103-115.

Chapter 2

Compulsory Voting, Turnout, and Government Spending: Evidence from Austria

Joint with Mitchell Hoffman (University of Toronto) and Gianmarco León (Universitat Pompeu Fabra and Barcelona Graduate School of Economics)

2.1 Introduction

Despite the centrality of elections to democracy, in elections around the world many people fail to vote. Many European countries have seen a steep decline in turnout rates in the past 30 years, with record low rates in the past two (2009 and 2014) elections for the European Parliament.[1] Ethnic minorities, immigrants, and poor voters in Europe are significantly less likely to vote, potentially distorting the political process (e.g., Gallego, 2007). In the US, turnout also exhibits large disparities along socioeconomic and racial lines (e.g., Linz et al., 2007; Timpone, 1998). Such disparities in turnout are believed to cause disadvantaged groups to be under-served by government (e.g., Meltzer and Richard, 1981; Lijphart, 1997).

One policy to help address these issues is to make voting mandatory. As of 2008, 32 countries had a compulsory voting (CV) law in place (Chong and Oliv-

[1]From http://www.europarl.europa.eu/elections2014-results/en/turnout.html, last accessed March 16, 2016.

era, 2008), and a higher number had CV at some point during the last 50 years. In March 2015, US President Barack Obama proposed the possibility of CV, arguing "If everyone voted, then it would completely change the political map of this country. The people who tend not to vote are young, they're lower income, they're skewed more heavily towards immigrant groups and minority groups...There's a reason why some folks try to keep them away from the polls."[2] However, there is limited empirical evidence about how CV affects turnout, party vote shares, and government policy.

We provide empirical evidence on the impact of CV laws on turnout, electoral outcomes, and fiscal policy using a unique natural experiment in Austria. Since World War II, Austria's nine states have changed their CV laws at different times for different types of elections. Austria provides a compelling case study for multiple reasons. First, the variation in CV laws is significant across states and over time, providing rich variation for quasi-experimental analysis. Second, like the US and many other countries, Austria exhibits socioeconomic disparities in turnout, with poor people being less likely to vote than the rich (Mahler et al., 2014). In addition, as noted by Ferwerda (2014), with the exception of one Swiss canton (Vaud), Austria is the sole modern democracy to have within-country variation in CV for national elections.

Using state-level voting records on state and national elections from 1949-2010, we find that CV increases turnout from roughly 80% to 90%. Impacts on turnout vary somewhat across the three types of elections (parliamentary, state, and presidential), but are sizable. Interestingly, however, CV does not appear to affect state-level spending. These "zero effects" are reasonably precisely estimated and robust to different specifications that deal with concerns regarding possible endogenous changes in CV laws.

How could it be that CV had large impacts on turnout, but did not affect policy outcomes? Our analysis shows that despite the large increase in turnout, CV did not affect electoral outcomes: vote shares for liberal parties did not change significantly, nor did the number of parties running for office or the victory margin in state or parliamentary elections.

[2]See, e.g., http://www.cnn.com/2015/03/19/politics/obama-mandatory-voting/, accessed March 16, 2016.

To complement our main aggregate analysis and dig further into mechanisms, we use repeated cross sections of individual level data to analyze interaction effects of CV laws with voter characteristics. While our statistical power is more limited compared to our main analyses, we find suggestive evidence of larger CV impacts on turnout among women and those with lower income. Impacts also seem larger among those who have low interest in politics, who have no party affiliation, and who are relatively uninformed (as proxied by newspaper reading). While suggestive, these results are consistent with a story where voters who vote or abstain due to the introduction or repeal of CV do not have strong policy or partisan preferences (on average), thereby having little or no effect on electoral outcomes. Furthermore, if such voters are unresponsive to policy in deciding which party to support, parties may have little incentive to shape policies to suit those voters' preferences.

Our paper relates to three main literatures. First, an important literature analyzes how changes in turnout and electorate composition affect public policy (Persson and Tabellini, 2000), often looking at the impacts of enfranchising particular groups of people. For example, the enfranchisement of women in the US led to increases in government health expenditures (Miller, 2008), as did the adoption of electronic voting in Brazil, which effectively enfranchised illiterate voters (Fujiwara, 2015). Similarly, Naidu (2012) shows that post-Civil War laws restricting voting for blacks in the US South had sizable impacts on public policy.[3] Our findings do not contradict this literature, but complement it, suggesting that the extent to which changes in turnout affect policy depends importantly on whether these policies affect a group of the population with specific policy preferences.

Second, our paper speaks to the literature on the determinants of voter turnout. As discussed in Gerber and Green (2012), scholars have analyzed numerous interventions aimed at increasing turnout, often using randomized experiments. In non-experimental studies, a significant literature examines the impact of voting

[3]Other papers in this literature show mixed results of the extension of the voting franchise on redistributive policies (e.g., Husted and Kenny, 1997; Rodriguez, 1999; Gradstein and Milanovic, 2004; Timpone, 2005; Cascio and Washington, 2014). A common message from this literature is that efforts to extend the voting franchise can significantly affect public policy, making it more aligned with voters' preferences. Most of this literature analyzes episodes in which groups with specific policy preferences are *de jure* or *de facto* enfranchised, leading elected officials to cater policies toward them.

costs, often reaching different results from different changes in costs.[4] For example, Farber (2009) shows that election holidays and "time-off" have little impact on turnout in the US, whereas Brady and McNulty (2011) show that an increase in voting costs (due to unexpected changes in the location of polling stations) reduces turnout. We complement this literature by simultaneously analyzing turnout and government policy.

Two noteworthy recent studies follow this tradition, analyzing how changes in voting costs affect turnout and policy outcomes. Hodler et al. (2015) propose a model of government where higher-skill individuals are more likely to vote. A reduction in voting costs leads to some lower-skill individuals choosing to vote. On one hand, these lower-skill individuals like government spending because a greater share of it is paid for by the rich (Meltzer and Richard, 1981). On the other hand, they invest less in political information than high-skill voters, making them more likely to be impressionable and thereby driven to choose candidates based on lobby-funded campaign spending instead of government spending. To test the model, they study the staggered introduction of postal voting across Swiss cantons. Postal voting led to increased turnout, lower education of participants, lower political information, and lower welfare spending. Godefroy and Henry (2015) analyze the impact of voting cost shocks on the selection of politicians and discretionary expenditures. Using digestive infections as a shock to voting costs, they find that unanticipated increases in voting costs lead to lower turnout, higher candidate quality, and higher infrastructure expenditures in French cities. We discuss differences between our results and these two studies in Section 2.4.2.

Third, it relates to a small but burgeoning literature analyzing CV. Among a number of theory papers, Börgers (2004) and Krishna and Morgan (2011) argue that CV reduces welfare, whereas Krasa and Polborn (2009) show that compulsory voting (or costly voting) allows an aggregation of preferences that can increase welfare. In empirical work, Funk (2007) finds that abolishing CV significantly

[4]Weather shocks have been used as exogenous shifts in the cost of voting (e.g., Knack, 1994; Gomez et al., 2007; Hansford and Gomez, 2010; Fraga and Hersh, 2010; Gomez et al., 2007), as have general rules of governance (Hinnerich and Pettersson-Lidbom, 2014; Herrera et al., 2014), candidates' ethnicity (Washington, 2006), and availability of certain information technology (Stromberg, 2004; Enikolopov et al., 2010; Gentzkow, 2006; Gentzkow et al., 2011; Gavazza et al., 2014). Some of these voting cost shifters are unexpected shocks, whereas others could be anticipated by politicians.

decreased turnout in Switzerland despite the fact that fines were small and not enforced. Her results highlight the expressive value of CV, an interpretation that could also apply to our setting, given the low levels of enforcement of the fines. However, this study does not investigate further the effects of changes in turnout on public policy.[5] In a cross-country study, Chong and Olivera (2008) show that countries with CV have lower income inequality. Fowler (2013) exploits the staggered introduction of CV across Australian states, finding that CV led to large increases in turnout. De Leon and Rizzi (2014) analyze students in Brazil, where voting is voluntary between ages 16 to 18, but mandatory afterward. They find that CV increases turnout, but does not affect political information. Using a field experiment in Peru providing information about changes in abstention fines, León (2017) shows that a reduction in the fines decreases turnout, and consistent with our findings, that the reduction is driven by uninformed, uninterested, and centrist voters. However, León (2017) can only analyze *policy preferences* and cannot analyze *actual policies* as our paper does.

A few political science papers involve the specific case of CV in Austria. The first paper to explore it was Hirczy (1994), who compared overall voting rates between Austrian states over time; the graphical evidence presented suggests that adoption of CV led to significant increases in turnout. The paper closest (and contemporaneous) to ours is Ferwerda (2014), who analyzes the effects of the repeal of CV by the Austrian parliament in 1992 on turnout and on changes in party vote shares. Although his analysis period is much shorter, the magnitude of the effects found on electoral participation and party vote shares are broadly consistent with ours.[6,7] Our paper goes beyond these studies in three main ways. First and foremost, not only do we analyze the political consequences of CV, but also impacts on spending, thereby providing the first micro study (for Austria or any other

[5] Also focusing on Switzerland, in recent work contemporaneous with our own, Bechtel et al. (2016b) find that CV increases electoral support in referendums for leftist policy positions. Note, however, that Bechtel et al. (2016b) study voters' support for policies as opposed to the behavior/policies of representative government, which we study. Still using Switzerland, Bechtel et al. (2016a) find that CV does not lead to habit formation in voting.

[6] Ferwerda (2014) also uses municipal-level data instead of state-level data.

[7] Another contemporaneous paper, Shineman (2014), also uses Austria as a case study to demonstrate the effects of CV on individual-level political sophistication, finding that both recent and long-term exposure to CV increase voters' information.

country) to examine how CV affects government spending. Second, we complement the analysis of aggregate data with individual level information on political preferences and voting behavior, allowing us to study the shift in the composition of the pool of voters resulting from CV. Finally, we analyze all elections from 1949-2010 instead of just a subset; this enables us to implement a fixed effects analysis allowing for different state linear trends, ruling out the concern that the effects are only valid in the short term and that we should expect a reversion to the mean.

Section 2.2 provides background on democratic institutions and CV in Austria. Section 2.3 describes the data. Section 2.4 discusses our estimation strategy and shows the results. Section 2.5 analyzes mechanisms for our results. Section 2.6 concludes and discusses external validity.

2.2 Institutional Background

2.2.1 Democratic Institutions and Budgeting Processes in Austria

Austria is a federal and parliamentary democracy, composed of nine autonomous states. The National Parliament is composed of two chambers, the National Council (*Nationalrat*) and the Federal Council (*Bundesrat*), with legislative authority vested mostly in the former. National Council members are directly elected by proportional representation, whereas members of the Federal Council are elected by the state legislatures. Austria's executive branch is composed of the Federal President (*Bundespräsident*), the Federal Chancellor (*Bundeskanzler*), and the Federal Cabinet. The Federal President is elected by simple majority in a popular election, and the candidates are nominated by party coalitions. The president holds the mostly ceremonial position of head of state. The Federal Cabinet is composed of the Federal Chancellor (the head of government) and a group of ministers, all of whom are appointed by the president. Austrian states are ruled by their own regional parliament (*Landtag*), a state government (*Landesregierung*), and a governor (*Landeshauptmann*). State parliament representatives are directly elected. Unlike the federal government, state governors are elected by the state parliament.

Over 95% of taxes in Austria are collected at the federal level (Blöndal and Bergvall, 2007). Taxes are distributed across the three levels of government (federal, state, and local) according to Fiscal Equalization Laws, which last for short periods of time (~ 4 years) and are established by a consensus between the federal and regional governments (Blöndal and Bergvall, 2007). Within the two lower levels of government, tax revenues are distributed across the different units according to a formula, which takes into account demographic and revenue criteria. Federal transfers to state governments are classified into two broad categories: (i) funds earmarked for a precise purpose and (ii) discretionary funds (1948 Constitutional Law, Sections II and III).[8] Throughout our period of analysis, discretionary funds tended to account for about half of the total transfers (Lehner, 1997), giving state governments considerable fiscal autonomy. States' spending autonomy comes across in the substantial variation in state spending percentages on different budget categories.[9] Although the largest portion of tax revenues are allocated to the central government, state governments receive a significant portion of the total budget, and are responsible for providing a wide array of public goods and services. In 2006, for example, spending by state governments accounted for 17% of total spending, with 70% and 13% of spending carried out by the central and municipal governments, respectively (IMF, 2008). State government responsibilities include primary education, regional infrastructure, hospitals, transportation, social welfare, and pensions for state civil servants (IMF, 2008).[10]

In the postwar period, there were two major parties in Austria, as well as

[8] We use "discretionary" to refer to non-earmarked transfers. From the discussion in Lehner (1997), much of the earmarked funds are meant to be spent on wages, particularly wages of teachers. Earmarked funds are also used for infrastructure expenses, housing subsidies, residential dwelling projects, local transport, disaster control, environmental/agricultural expenses, and health.

[9] During 1980-2012, for example, the government of Burgenland devoted 66% of its budget to welfare expenditures and only 13% to infrastructure spending, whereas the neighboring state of Lower Austria spent 43% of its resources on welfare, and 40% on infrastructure. Section 2.3 describes the budget categories further.

[10] In some of these areas, the responsibilities of the central and state governments overlap and are thus co-financed or managed jointly (IMF, 2008). There is also overlap of responsibilities between state and municipal governments (IMF, 2008). OECD (1997) notes for Austria that "the revenue-sharing process is rather complex" (pp. 99-100). For further details on state spending, see 2.9.

several smaller parties. The two major parties were the center-right People's Party (ÖVP) and the center-left Social Democratic Party (SPÖ). In addition, there were the right-wing populist Freedom Party of Austria (FPÖ), the Communist Party of Austria (KPÖ), and the Green Party, as well as the Alliance for the Future of Austria, the Liberal Forum, and others.

2.2.2 Compulsory Voting in Austria

Figure 2.1 summarizes the process by which CV was introduced and later repealed in Austria. The mandate to vote was changed a number times during 1949-2010; whether voting was compulsory varied substantially both across and within states, and depending on the type of election, as seen in Figure 2.2. CV was first introduced in Austria in the 1929 Constitution. In particular, voting became mandatory for all citizens in presidential elections, but it was up to each state to determine whether voting was mandatory or voluntary in parliamentary and state elections (see 2.8.1 for further details).

The first presidential election with CV was held in 1951. Up until 1980, there were seven presidential elections, and all of them had CV. However, a 1982 amendment to the Austrian Constitution made voting in presidential elections compulsory only in the states that decided so. In the 1986 presidential elections, the states of Vorarlberg, Tyrol, Styria, and Carinthia decided to keep CV. Furthermore, Carinthia enacted a law establishing CV for parliamentary and state elections. The remaining five states abolished CV in presidential elections after the 1982 amendment.

In 1992, a Federal Constitution amendment by the national parliament withdrew the power of establishing mandatory voting in the national parliament elections from the states (Federal Law Gazette No. 470/1992). Starting in the 1994 parliamentary elections, voting was optional in all states. After this constitutional amendment, the states which still had CV in presidential and state parliament elections started repealing their state laws one by one. In 1993, Carinthia and Styria eliminated CV for both types of elections. Tyrol repealed CV for state parliament elections in 2002, and Vorarlberg got rid of it before the 2004 elections. After these elections, Tyrol finally repealed CV for presidential elections. Thus, the

2010 presidential elections (the last in our sample) were the first in which voting was voluntary throughout the country.

During the period when voting was compulsory, local authorities were responsible for issuing fines against non-voters failing to provide a reasonable excuse for abstaining.[11] Abstention penalties were extremely rare, as the law allowed for a wide range of excuses for not voting, such as illnesses, professional commitments, urgent family matters, being outside the state during the election, or "other compelling circumstances" due to which the voter could not go to the polls. Importantly, voters who excused themselves were not required to provide documentation justifying their absence.[12] After qualitative work with Austrian citizens and elites, Shineman (2014) concludes that fines rarely had real consequences and were almost never enforced. 2.8.2 gives additional details supporting that fines were weakly enforced.

2.3 Data Sources and Descriptive Statistics

In the empirical analysis, we draw upon three main sources of information. First, to analyze how CV affects turnout and political competition, we collected election data. Our initial sample consists of all parliamentary, presidential, and state elections held since the end of World War II until 2010.[13] For these elections, we hand-collected data on voter turnout, invalid ballots and party vote shares from the Austrian Federal Ministry of the Interior's yearbooks. We define turnout as the share of registered voters who show up to the polls.[14]

[11]Federal Presidential Election Law, Article 23 and Federal Parliament Election Law, Article 105 (4). We provide details on the maximum fine amounts specified in the law and their evolution in different states in 2.8.2. While there is information on maximum fine amounts in states, we have limited information on actual fine amounts (for the cases where fines were actually enforced), due to the involvement of local authorities in setting actual fines.

[12]Abstainers needed to provide evidence of reasons for not voting only when an administrative penal procedure was initiated against them. 2.8.2 provides further details on abstention sanctions.

[13]We exclude the 1945 election. This election was unusual in multiple respects, coming just after World War II and banning former Nazis (about one-tenth of the voting age population) who were not allowed to vote until the 1949 election (Bischof and Plasser, 2008). In the period under consideration, there were 19 parliamentary elections, 12 presidential elections, and around 11-15 state elections in each of the nine Austrian states.

[14]As discussed by Geys (2006), studies of turnout have used different denominators in defining turnout, including voting age population, eligible voters, and registered voters. Geys (2006) argues

Table 2.1 gives descriptive statistics. Average turnout in our sample is relatively high, ranging from 86% in state elections to 90% in parliamentary elections. The average incidence of invalid ballots in these elections is 2%. We define "right-wing" parties as ÖVP and FPÖ and "left-wing" parties as SPÖ and KPÖ.[15] Both in state and parliamentary elections, the right-wing vote share (52%-53%) exceeds the left-wing vote share (around 40%).

Second, to analyze spending, we draw upon detailed annual information on expenditures by each of the state governments, which is publicly available on the Austrian Statistical Agency's website. Unfortunately, this information is only available since 1980.[16] The Austrian Statistical Agency expenditure data include 10 expenditure groupings by year. To simplify exposition and improve statistical precision, we combined the groupings into three broad categories: Administrative, Welfare, and Infrastructure. We define "administrative expenditures" as spending on elected representatives and general administration. "Welfare Expenditures" comprise expenditures on education; health; arts and culture; and social welfare and housing. "Infrastructure Expenditures" are those for construction, transport, and security.[17] The yearly expenditure data is expressed in millions of 2010 euros. In the 1980-2012 period, a majority of expenses (54%) were devoted to the social sector, while 25% of all resources were spent on administration, and the remaining 21% were devoted to infrastructure. In all of our main specifications, we also include state-specific, time-varying covariates (i.e., total population and unemployment rates) obtained from the Austrian Statistical Agency.

that there is no single correct measure of turnout, but that using eligible voters is preferred over voting age population when data on eligible voters are available, since it excludes those who are legally forbidden from voting (Geys, 2006, p. 639). For Austria, registration is automatic for all citizens with a permanent residence in the country; thus, "registered voters" and "eligible voters" are approximately equivalent in Austria. In addition, we prefer using registered voters over voting age population because we were only able to gather state-specific voting age population data beginning in 1982. Finally, an advantage of using registered voters is that the data are based on administrative registration records.

[15] 2.9 gives further discussion and details on this grouping.

[16] This restricts our analysis to 10 parliamentary elections, 6 presidential elections and 6-7 state elections in each state. In Appendix Table 2.A3, we repeat our main results on turnout, invalid votes and political competition for the restricted period of 1979-2010, and they are qualitatively similar to those in the main text covering 1949-2010.

[17] Table 2.A1 provides a detailed description of expenditure areas falling into each of the 10 groupings. Our conclusions are substantively unchanged if we analyze the 10 groupings individually.

Third, to analyze heterogeneity in CV turnout effects according to individual characteristics, we use the Austrian Social Survey (ASS), a nationally representative survey conducted in 1986, 1993, and 2003. The 1993 round did not include questions on turnout, so we exclude it from our analysis. Haller et al. (1987) and Haller et al. (2005) give details on the 1986 and 2003 rounds. The ASS asks respondents standard questions on demographics, socioeconomic status, education, and importantly, it inquires about voting behavior, and political and social preferences. Table 2.2 shows summary statistics from our individual level data.[18]

2.3.1 Comparing Austria to Other Countries

Before providing our results, we seek to provide context by comparing Austria to other countries in terms of political behavior. Figure 2.3 compares turnout rates in Austria and other OECD countries. While Austria has high turnout, it is not an extreme outlier and there are a number of other countries with broadly similar turnout levels. While the median turnout in this sample is 75.7%, turnout in Austrian parliamentary (state) elections when voting was voluntary is 83.8% (77.5%), which places these elections in the 76^{th} (56^{th}) percentile of the turnout distribution.[19]

Table 2.A2 compares Austria to other rich countries on three non-turnout measures of political engagement: interest in politics, information acquisition (prox-

[18]Our sample restricts to individuals (i) who were of voting age in the previous parliamentary elections and (ii) who reported whether or not they voted in the previous parliamentary elections. Only 3% of people failed to provide information about whether they voted, and missingness is not correlated with whether there is CV in their state. In the 1980s, people were allowed to vote if they were 19 or older, while the voting age was lowered to 18 in 1992 (Kritzinger et al., 2013). For analysis of the ASS data, we do not restrict the sample based on voter eligibility because information on voter eligibility is absent from the 2003 wave. Thus, the turnout denominator (voting age population, which is 19 or older in the 1986 round and 18 or older in the 2003 round) in our analysis of the ASS data (presented later in Table 2.6) differs slightly from the one used in our main turnout results (Table 2.3). Our main focus in the heterogeneity analysis is on differences across subgroups, so this should not be a central concern. Further, the Table 2.3 results are similar whether registered voters or voting age population is used as the denominator.

[19]The percentile numbers are calculated relative to elections with voluntary voting between 1979 and 2010 for the OECD countries in Figure 2.3. Being in the 76th percentile means that average turnout in Austrian parliamentary elections is higher than turnout in 76% of OECD elections. The state election percentile should be taken with some care because it reflects a comparison of state elections in Austria to national elections in other countries. In Figure 2.3, turnout is defined as a percentage of registered voters.

ied by newspaper reading), and political party membership. Austrians are politically engaged, but seem broadly similar to other rich countries. In the 2003 ASS, 26% of Austrians reported being "Very Interested in Politics," which is comparable with Switzerland (26.6%) or Germany (21%), but slightly higher than other OECD countries participating in the World Value Survey.[20] While 68.7% of Austrians in the ASS read the newspaper regularly, 74.8% of people from other OECD countries in the World Value Survey report having read the newspaper the week before they were surveyed. The level of information acquisition in Austria is below countries like Switzerland (91.3%), Sweden (94.5%), or Japan (88.8%), and only above less developed democracies like Hungary (56.8%), Poland (55.1%), or Spain (62.7%). Likewise, 12.4% of Austrians are members of a political party, comparable to 13.4% of respondents in other OECD countries. These statistics provide reassurance that our results seem unlikely to be due to an unusual institutional context or by political behaviors that are highly specific to Austria.

2.4 Empirical Strategy and Results

We estimate the effect of CV laws (in different elections) on turnout, invalid ballots, electoral outcomes, and state-level public spending. Using a difference-in-differences model, we compare states with and without CV at different points in time. Our baseline specification is:

$$y_{st} = \alpha_0 + \beta_1 CV_{st} + X_{st}\beta_2 + \delta_s + \nu_t + \gamma_{st} + \epsilon_{st} \tag{2.1}$$

where y_{st} is an outcome variable in state s and year/election t;[21] CV_{st} is a dummy

[20]Note that Austria did not participate in the WVS, so the numbers are not strictly comparable, but they do give us a broad sense of how Austrians' political opinions compare to other countries.

[21]For the case of government expenditures, the timing of the dependent variable is slightly different. We analyze spending as a function of CV in the current "electoral period" with a one year lag. We define an electoral period as lasting from the year of an election until the year before the next election. Thus, for example, if elections takes place in years t and $t + 4$, we consider expenditures in years $t + 1$ to $t + 4$ as a function of CV in year t. The idea is that the potential impact of CV on spending may come through elections, and we want to allow an election to take place in order to allow the possible consequences of CV on spending to occur. Also, for our spending analysis, the state-specific trends are by year instead of by election.

for whether voting was compulsory in year/election t and state s; X_{st} is a vector of state-year covariates (population and the unemployment rate); δ_s are state fixed effects; ν_t are year fixed effects; γ_{st} are state-specific linear trends (at the election level); and ϵ_{st} is an error. We run these regressions separately for different types of elections (parliamentary, state, and presidential). We also do a pooled specification including the three types of elections together—when we do this, we also include election type dummies, as well as all three sets of state-specific trends.

We allow for arbitrary within-state correlation of the errors by clustering our standard errors at the state level. Given the small number of clusters, our standard errors might be inconsistently estimated (Bertrand et al., 2004). Following Cameron et al. (2008), we also report wild-bootstrap p-values.[22]

In these specifications, our "treatment group" are statesXelectoral periods subject to CV, while the "control group" comprises those in which there is voluntary voting (VV). At any given period in time, we compare states under CV vs. VV (leveraging the time fixed effects), and at the same time, we make within-state comparisons, comparing electoral terms with and without CV (using state fixed effects). Using state level data, we analyze the effect of CV on: (i) turnout and valid ballots; (ii) left/right vote shares, number of parties, vote shares and margin of victory of the winning party; and (iii) government expenditures in social services, administration, and infrastructure. For (i) and (ii), the unit of analysis is stateXelection; when analyzing the impact of CV on expenditures, the unit of analysis is a stateXyear.

2.4.1 Turnout and Invalid Votes

Even with weak enforcement, as is the case for Austria, CV can affect turnout through the signaling value of enacting a law, as argued in Funk (2007). Panel A in Table 2.3 shows the effects of CV on turnout within and across Austrian states in the 1949-2010 period. The introduction of CV causes statistically and economically significant increases in turnout in parliamentary, state, and presidential elections.

[22] We calculated the wild-bootstrap p-values using the *cgmwildboot* program created by Judson Caskey, and imposed the null hypothesis, as recommended by Cameron et al. (2008).

When independently considering each type of election, we find that CV increases turnout by 6.5 percentage points in parliamentary elections, by 17.2 percentage points in state elections, and by 9.5 percentage points in presidential elections. However, we gain additional power by pooling all types of elections together, as doing so allows more precise estimation of the year and state fixed effects. In column 4 of Panel A in Table 2.3, we pool the three types of elections together, and analyze the impact of CV on each type of election (our preferred specification). CV now increases turnout by 6.6, 8.1, and 9.1 percentage points for parliamentary, state, and presidential elections, respectively. Note that these results show slightly lower point estimates than in the previous regressions, and this is particularly the case for state elections, for which we have a smaller sample size. The results are highly significant based on standard errors clustered by state (in parentheses) or based on wild bootstrap p-values (in brackets).[23]

CV can increase turnout by drawing uninterested voters, or those who might not be familiar with the voting process. If this is the case, we might expect the proportion of invalid ballots to rise. As shown in Panel B of Table 2.3, the increase in turnout from CV is paired with a statistically significant increase in invalid votes. In elections without CV, the share of invalid votes ranges between 1.5% and 3.8%. Based on the results in the preferred specification (column 4), CV increases the share of invalid votes by 0.9–1.8 percentage points, depending on the type of election. Even though the increase in turnout associated with CV is also conducive to a higher proportion of invalid votes, there is certainly not a one-to-one relation. That is, for every 10 people who are driven to vote due to CV, only 1.5–3 of them issue an invalid ballot, while the others correctly vote for a party or candidate. Hence, an increase in turnout of this magnitude could very well result in a shift in election results and public policies.[24]

[23] Note that the wild bootstrap procedure does not deliver bootstrapped standard errors, but rather p-values. The p-values found using the clustered standard errors and the wild bootstrap procedure are generally similar in most cases. Throughout the paper, for ease of exposition, our interpretation of confidence intervals is based on the clustered standard errors.

[24] Given that the analysis of the effects of CV on fiscal behavior is performed only for those years for which expenditure data is available (1979–2010), we re-run the analysis from Table 2.3 on a comparable sample in Appendix table 2.A3. The results shown for turnout and invalid votes are comparable to the ones for 1949–2012.

2.4.2 Public Spending

An increase in participation rates could potentially affect government spending in multiple ways. If preferences for public goods in the participating electorate are now different, the government might change the distribution of public spending, keeping the size of the overall budget constant, but shifting it between sectors. Alternatively, the overall size of the budget could change by pushing the local government or local parliamentarians to change deficits or to negotiate larger budgets from the federal government.

As discussed in Section 2.2.1, given the ceremonial role of the federal president in Austria, we do not expect to see effects of CV in presidential elections on spending—as a placebo test, we run regressions estimating the effect of CV in presidential elections on fiscal behavior at the state level, and as expected, we do not find any effects (Appendix Table 2.A4). On the other hand, the national parliament decides on the resources that each state government gets. In addition, state parliaments are in charge of preparing the state's budget, and thus laws that affect this level of government may also affect spending. In this section, we turn our attention to the effects of CV on fiscal policy at the state level.

In the subsequent analysis, we study total state expenditures, as well as their composition: administrative, welfare, and infrastructure expenditures. For each spending category, we analyze three different measures of fiscal policy, which are intended to capture the different mechanisms described above: (i) the log levels, (ii) the log per-capita, and (iii) as a percentage of the total budget. We use a similar estimation framework as in Section 2.4.1.

Table 2.4 shows no consistent evidence of CV affecting the amount or composition of public spending. Most coefficients are close to zero, and the clustered standard errors as well as the wild-bootstrap p-values indicate that there is no significant relationship between CV and total budget or its composition. Across the 12 regressions in Panel A of Table 2.4, the coefficients are sometimes positive and sometimes negative, but small in magnitude. They are also relatively precise. For example, the estimated coefficient in column 1 on total spending corresponds to a 95% confidence interval of $[-0.009, 0.071]$, meaning we can reject that CV decreases total spending in more than 0.9% or that it increases it by more than about

7.1%. Similarly, the point estimates on administrative, welfare, and infrastructure spending are relatively close to zero, at 0.6%, 3.5%, and 6.6%, respectively.[25] In contrast, electronic voting in Brazil (Fujiwara, 2015, p.452) and US women's suffrage (Miller, 2008, p.1289) are estimated to have each raised health spending by about one-third.

Our result on public spending contrast with those of Hodler et al. (2015) and Godefroy and Henry (2015). Consistent with a model where decreases in voting costs increase the share of voters who are uninformed, Hodler et al. (2015) find that Swiss postal voting decreased welfare spending by 4-7% and business taxation by 3-7%. Why might our result differ? First, there are various political differences between Switzerland and Austria, including that Switzerland has much lower turnout (though other levels of political involvement and interest do not seem so different). For example, following the idea of the model of Hodler et al. (2015), it could be that there are few "impressionable" voters for politicians to take advantage of in Austria, and this could limit whether there are impacts on spending. Second, postal voting and CV may have different impacts on a political system. It is not clear that the population of voters who would respond to CV are the same as those would respond to postal voting. Even if both populations seem more uninformed on average, they may differ on unobservables. We note, though, that these explanations are speculative and cannot fully resolve why there are differences across papers. Further research is needed.[26]

Godefroy and Henry (2015) show that a decrease in turnout in French cities due to more digestive infections decreases harmful financial decisions, and leads to higher subsidies obtained by the municipality and more infrastructure expen-

[25] Besides using the 3 broad expenditure categories, we also do the analysis using the 10 more disaggregated groupings in Appendix Tables 2.A5 and 2.A6. Given the more granular level of the data, some of the confidence intervals are relatively large (e.g., for finance and service expenditures). However, the conclusions are qualitatively robust, with coefficients mostly statistically insignificant and close to zero. Further, our zero results are also qualitatively robust under an instrumental variable approach, where we estimate the effect of turnout (instrumented by CV) on expenditures (Appendix Table 2.A7).

[26] There are a number of other potentially significant political differences between Switzerland and Austria. These include that Swiss cantons seem more powerful than Austrian states (e.g., unlike Swiss cantons, Austrian states do not have significant tax powers (Fuentes et al., 2006)); that Austria is more linguistically homogeneous; and that Switzerland and Austria have different historical experiences with democracy (e.g., Austrian politics has often emphasized consensus (Kritzinger et al., 2013)).

diture, and they argue that this is due to the selection of better qualified politicians. In our paper, it is difficult to test this hypothesis directly because we do not observe direct information on the quality of politicians (such as whether a mayor contracted a toxic loan, which is observed by Godefroy and Henry (2015)). In addition, in contrast to us, Godefroy and Henry (2015) study unanticipated, unknown-in-advance shocks to voting cost.

2.4.3 Robustness Checks

The identification assumption in our main regressions is that CV is uncorrelated with unobserved time-varying state characteristics once we have controlled for time invariant, state-specific factors, as well as year-specific, state invariant factors, and partialled out state-specific linear time trends. For example, if conservative states are more likely to support CV, this should be absorbed by our state fixed effects. On the other hand, if there is a national push for abolishing these types of laws (e.g., in 1982), this would be captured by the year fixed effects. One threat to our identification assumption is that, even though some of the changes in CV laws were issued by the federal parliament (e.g., the 1992 repeal of CV in parliamentary elections), and thus are unlikely to respond to state-specific political dynamics, others changes were issued at the state level, and these decisions might be related to voting trends. As in any difference-in-differences model, this is the same as assuming that, conditional on the set of observables and fixed effects, the trends in voting, political competition, and expenditures in states in which CV was introduced were the same as in states where voluntary voting was in place; if the new voting regime had not been enacted, e.g., they have parallel trends in the pre-treatment period.

The parallel trends assumption would be violated if the states most likely to implement CV were those in which turnout was downward trending. In this case, an estimation relying on simple fixed effect will understate the effect of CV laws. Similarly, state governments might find it easier to enact CV laws when turnout is trending upward, since enforcement costs will be lower in these states. In this case, a fixed effects model would overestimate the results. The inclusion of state-specific time trends controls for any linear trends in our outcome variables, and thus partially addresses these concerns, but further tests are needed.

As mentioned in Section 2.2.2, in our study period there is one change in CV laws that is unrelated to any state–year specific characteristic, namely, the one introduced by the federal government in 1992.[27] This Federal Constitution amendment withdrew the prerogative of establishing mandatory voting in the national parliament elections from the states. Effectively, while some states already had voluntary voting in parliamentary elections, others (Vorarlberg, Styria, Tyrol and Carinthia) were forced to adopt it. Figure 2.4 shows the evolution of turnout, invalid votes, and total, administrative, welfare, and infrastructure expenditures in the same analysis period (1986-2011), for states that never had CV and those that were mandated to eliminate it in 1992. States that had CV before 1992 had higher turnout and more invalid ballots, but importantly, before CV is abolished, the trends in these variables run parallel to the ones in states that did not have CV before 1992. Similarly, in our four expenditure variables, for which we do not observe an effect of the elimination of CV, the trends for both types of states run parallel during the whole study period.[28]

To further alleviate the concern that CV laws might have been introduced responding to changes in our dependent variables of interest, we include leads and lags of CV in our regressions. If it were the case that CV laws responded to changes in turnout, we would expect turnout in period t to be correlated to either CV in $t + 1$ or CV in $t - 1$. The results for our preferred specification in Table 2.A9 show that, besides the contemporaneous effect of CV on turnout and invalid votes, the introduction of CV in the previous election or next electoral period has no effect on our variables of interest. The estimated effects for the three types of elections generally show a "zero" of the lags and leads of our independent variable, i.e., there were no pre-trends, or anticipation effects. (One exception is the surprising significance of the coefficient of the lead of CV for presidential elec-

[27]Ferwerda (2014) uses this federal change in legislation to explore changes in party vote shares; he argues that, given that it was issued at the federal level, it is independent of political dynamics at the local level.

[28]We also perform a difference-in-differences regression limiting our sample to the parliamentary elections in the electoral periods between 1986 and 2011. The magnitude and statistical significance of the results is similar to those shown in Tables 2.3 and 2.4, suggesting that any other changes in CV (besides the 1992 one) are unlikely to be correlated with trends in the main dependent variables. For further details turn to Table 2.A8 and 2.9.

tions.)

A potential concern is that authorities anticipate the introduction/repeal of CV laws and alter the level or composition of public spending before the law change takes place. If this were the case, we would observe a correlation between public spending in year t and CV in $t + 1$. Alternatively, any delays in the reaction of public spending to changes in CV laws would lead to a correlation between CV in $t - 1$ and public spending in year t, which would not be captured in our baseline specification. As seen in Table 2.A10, spending is uncorrelated with CV in the past, current, or future electoral period for all types of elections.

Together, Figure 2.4 and Tables 2.A8-2.A10 provide evidence supporting the parallel trend assumption, and help rule out potential reverse causation between turnout and CV. As an additional robustness check, in 2.9, we discuss heterogeneity in our spending (and turnout) results according to levels of turnout in different states.

2.5 Understanding the "Null Effect" on Policy Outcomes

How could it be that CV had sizable impacts on turnout, increasing the number of valid votes, but did not affect policy outcomes? One potential explanation for these results is that the political choices of people who turn out because of CV are, on average, similar to the ones of people who would have voted even in the absence of CV. Another explanation is that median voter preferences may have changed, potentially leading to changes in electoral outcomes, but government spending still does not change for some other reason (e.g., commitment or agency issues).[29]

Besides exploring electoral outcomes, we also attempt to shed light on identi-

[29] For example, in citizen candidate models (Besley and Coate, 1997; Osborne and Silvinski, 1996), politicians may implement preferred policies that may differ from those of the median voter. In empirical work, Lee et al. (2004) (building on the model of Alesina, 1988) show that exogenous changes in party electoral strength do not affect the voting patterns of US congressmen. Another possibility is that uninformed or uninterested voters may have different preferences, but may be less responsive to policy than other voters; in such a scenario, politicians may have little incentive to change their policies.

fying the voters affected by CV. Several recent studies analyze large increases in turnout due to *de jure* or *de facto* enfranchisement of specific groups of the electorate, such as women in Miller (2008); the poor and illiterate in Fujiwara (2015); and African-Americans in Naidu (2012). Unlike these studies, we do not necessarily have a strong prior that people who vote because of CV make significantly different political choices than those who vote even when voting is voluntary.

2.5.1 Electoral Outcomes

Table 2.5 examines whether CV affects various electoral outcomes. We estimate similar regressions as in Section 2.4.1, but use as dependent variables the percentage of votes to the left or right wing parties, the number of parties, the share of votes of the winning party and its margin of victory (i.e., the difference in vote share between the winning party and the runner-up). For both parliamentary and state elections, CV does not affect the share of votes going to the right or left parties.[30] Further, there is no response from the political supply: the number of parties remains constant at about 6.9 and 6 for parliamentary and state elections, respectively. Finally, the party that wins the election does not receive a significantly different proportion of votes under CV, compared to states and elections in which voting is voluntary.[31]

In a related paper, Ferwerda (2014) exploits the 1992 constitutional change to analyze the effects of political participation on electoral results. Using municipality level data from 1990 and 1994 (before and after the change), he finds modest impacts the change on electoral outcomes. He finds a slight increase in votes for the left-wing SPÖ, coupled with a slight decrease in votes for minor parties, but his results are broadly similar to ours.

Overall, in Table 2.5, CV does not significantly affect party vote shares, the

[30]We do not perform these regressions for presidential elections because parties do not run as separate entities in those races. Additionally, we find no effects on individual party vote shares or voter polarization. See 2.9 for details. Somewhat relatedly in the literature, Martinez and Gill (2005) examine how turnout affects Democrat vote share in US presidential elections.

[31]Results for the 1979-2010 period shown in Appendix Table 2.A11 are comparable to those in Table 2.5. We find relatively larger point estimates for the effects of CV on party vote shares, but they are not statistically significant. Only for the case of votes for the left wing parties in parliamentary elections do we find a statistically significant impact of CV; however, the effect of CV in this case is small.

number of parties, or margin of victory. Under models with commitment issues, we might have expected CV to affect electoral outcomes, but we do not observe that in the data.

2.5.2 Composition of the Electorate

We use individual data from two rounds of the ASS (1986 and 2003) to examine what type of voters were most affected by CV. The goal is to better understand the mechanisms underlying our main results. We have information on turnout in the previous parliamentary election (1983 and 2002), and exploit within and between state variation in CV introduced by the federal abolition of CV between the surveys (in 1992). While no states had CV in the 2002 parliamentary election, 3 states (Styria, Tyrol, and Vorarlberg) had it in the 1983 elections. Our dependent variable is whether an individual voted in the previous parliamentary election, and our main regressor is a dummy for whether voting was compulsory in that election in the state where the respondent lives. We control for a set of individual covariates, as well as state and survey year fixed effects.[32] To examine what type of voters are more likely to respond to CV, we interact CV with various individual characteristics.[33]

Table 2.6 shows 8 separate regressions. In the left panel, we observe that impacts of CV seem to be larger among females, those with a vocational middle school degree, and those in the lower two income quartiles.[34] In this table, the

[32]Controls included in all regressions in this section include: age, age squared, gender, educational attainment, parents' education, working status, household size, community size, party preferences, party membership, interest in politics, and information acquisition (read newspaper regularly).

[33]Although the ASS asks about which party respondents voted for, we do not analyze this outcome because 17% of our sample did not answer this question, and attrition is differential along individuals' self-reported political preferences.

[34]In regression 1, we examine the impact of CV on turnout without interaction effects. CV increases turnout by 5.5 percentage points, slightly lower than the effect in the aggregate, state-level data. We must bear in mind, however, that these regressions rely on self-reported data, which might measure turnout with error. A standard concern could be that people might lie about whether they voted. We take comfort from the fact that the self-reported voting shares of 92% and 82% in 1983 and 2002 are relatively close to the actual voting shares in our data (93% and 85%, respectively). The voting shares are not directly comparable across datasets because our administrative turnout measure uses registered voters as the denominator, whereas the survey implicitly uses voting age population as the denominator (see footnote 18), but the general similarity suggests that

wild bootstrap p-values are larger than the implied p-values from the clustered standard errors, making the inference here more suggestive. In the right panel, we observe that the impact of CV is larger among people who are not party members, who are not informed (proxied by newspaper reading), who declare themselves uninterested in politics, and who report no party preference. For example, the coefficient of 0.148 in regression 8 indicates that CV increases turnout by 14.8 percentage points among individuals with no party preference. In contrast, for individuals who declare a party preference toward the main left or right parties in Austria, CV increases turnout by 3-4 percentage points.

Although CV may have affected the gender, educational attainment, and income of the median voter, our results indicate that it may not have shifted the median voter's political preferences. If those induced to vote by CV do not have strong political views, then such voters may not necessarily vote differently from the median voter in elections without CV. If the choices of the median voter are not shaped by CV (and if party platforms don't change), it would not be surprising if CV failed to affect what party wins and what policies are implemented. These results may also be consistent with a citizen candidate framework in which significant changes in the electorate are not necessarily accompanied by changes in policy outcomes. Unfortunately, we do not have enough data to precisely determine which of these two frameworks are more suited to explain the observed results.

2.6 External Validity and Conclusion

Although compulsory voting (CV) is often viewed as a way to foster voter turnout and consequently improve the representativeness of political processes, relatively little is known about how CV causally affects voter participation and, in particular, how it affects economic policy. We analyze these impacts by leveraging quasi-experimental variation in CV laws across Austria's nine states. We find that CV increased turnout from roughly 80% to 90%. This occurred even though penalties for not voting were rarely enforced. However, in our main results, the

---

misrepresentation seems unlikely to be a central issue in the data.

increase in turnout did not affect state-level spending (either in levels or shares of sectoral spending) or electoral outcomes. Effects of CV on turnout seem larger among individuals who are uninterested in politics, who do not have strong political views, and who are relatively uninformed (with informedness proxied by newspaper reading). This suggests that individuals swayed to vote by CV are more likely to exhibit these characteristics (compared to individuals not swayed by CV).

We view our results as consistent with a story where voters swayed to vote by CV do not cast different votes (on average) from those who vote regardless of the law. Our results may also be consistent with other explanations (e.g., citizen candidate models) where politicians implement policies that may not correspond to those preferred by the median voter. Ultimately, it is difficult to say definitively what theoretical mechanism explains our results. Our contribution, though, is to provide causal evidence (previously lacking) that CV laws need not significantly affect government spending. We believe this is important evidence for the policy debate regarding CV.

Overall, our results complement the literature documenting that extension of the voting franchise to specific population groups impacts policy. Our study suggests that policies aimed at increasing turnout (e.g., get-out-the-vote campaigns) need not necessarily affect public spending, and this seems particularly the case if these policies do not increase turnout among voters with specific policy preferences.

Our results are specific to Austria, so it is important to consider to what extent we think the results would extrapolate to other countries. As discussed earlier in Section 2.3.1, Austria has had relatively high turnout and political involvement even when CV was not in place, at least relative to the US and the OECD averages. At the same time, however, there are a sizable number of countries (e.g., Germany, the Scandinavian countries, and others) that share these features, particularly in Europe. Thus, we believe our findings may be relevant for these countries and other advanced democracies where reforms to increase political participation (such as CV laws) are being evaluated.

How might our results extrapolate for countries with lower levels of political involvement? In terms of turnout, one might imagine that countries with low initial turnout levels might experience even greater turnout increases (in absolute percentage levels) as a result of CV compared to countries like Austria with traditionally high turnout. Thus, our turnout findings could form a “lower bound” for impacts for reforms implemented in countries with low turnout.[35] Turning to government spending, as observed earlier in the paper, many countries without CV (such as the US) experience significant disparities in turnout along socio-economic lines, which are also correlated with levels of political interest. For Austria, our results suggest that CV induces low-interest or low-knowledge voters to participate. For countries with low initial turnout, while it is possible that voters induced to participate because of CV would have low interest in politics, it is also possible that a broader set of voters would be affected. This might cause CV to actually have a significant impact on government spending. However, as noted by Hodler et al. (2015), changes in voting costs could affect government spending in either direction, and it is challenging to make confident empirical predictions about possible impacts of CV in a setting very different from Austria. Thus, we urge significant caution in assessing the relevance of our results for countries like the US where turnout is much lower than in Austria.

Beyond political involvement, there are a host of other factors that may affect how CV affects government spending including whether there is a presidential or parliamentary system, and whether the population is relatively ethnically heterogeneous or homogenous. While we do not have strong priors on how such factors would influence the impact of CV, we cannot rule out that they may be at play. While our results provide the first quasi-experimental evidence on how CV affects government spending, they are certainly not the last word. We look forward to future research on how CV affects government spending, hopefully using data from additional countries.

[35] However, one can also imagine situations where CV laws have smaller effects on turnout. For example, countries with low turnout may be reflective of citizens being generally distrustful of government and authority. Such characteristics could lead to CV laws having less of an effect on turnout.

Figure 2.1: Evolution of CV Laws (1918-2010)

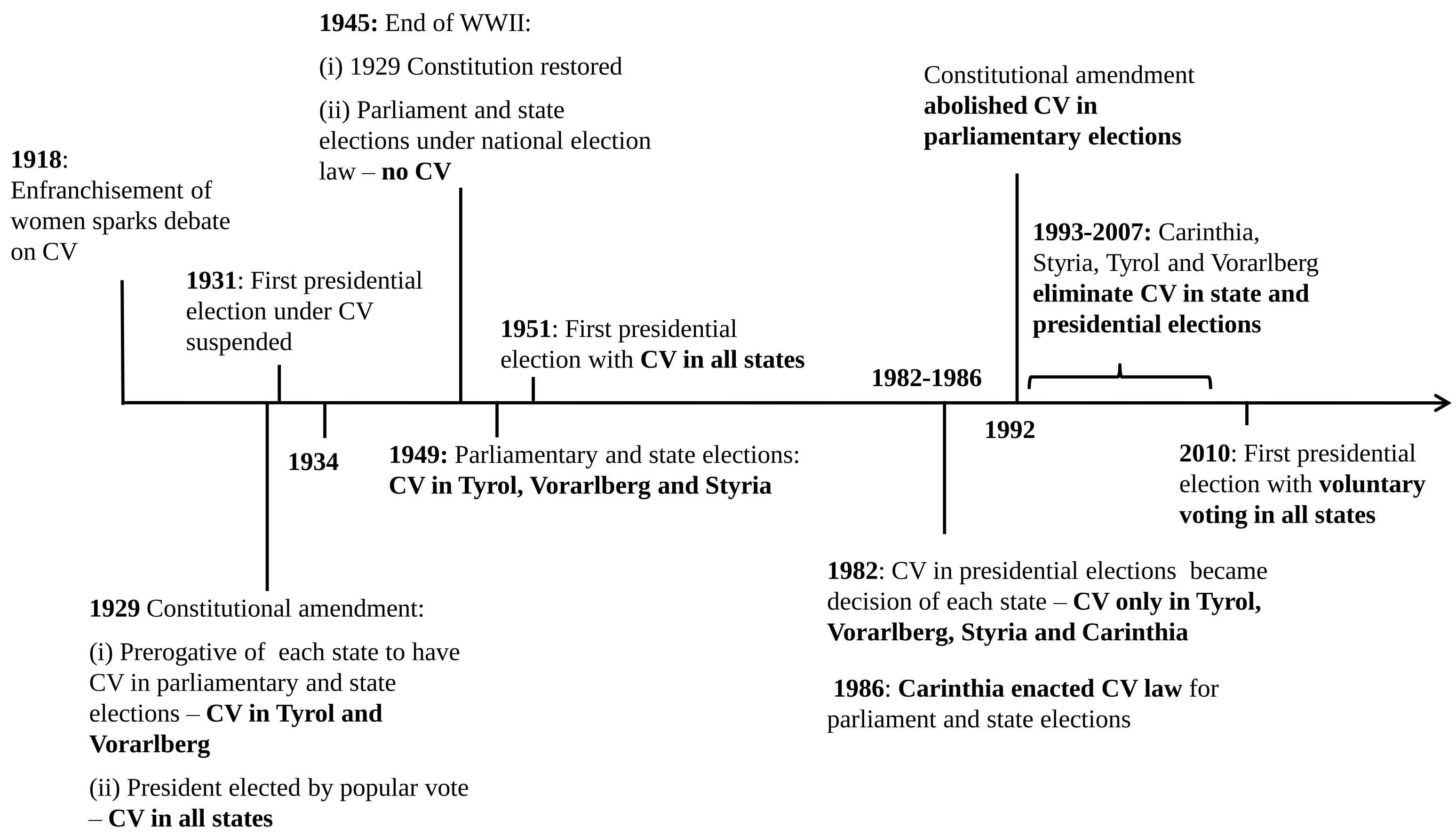
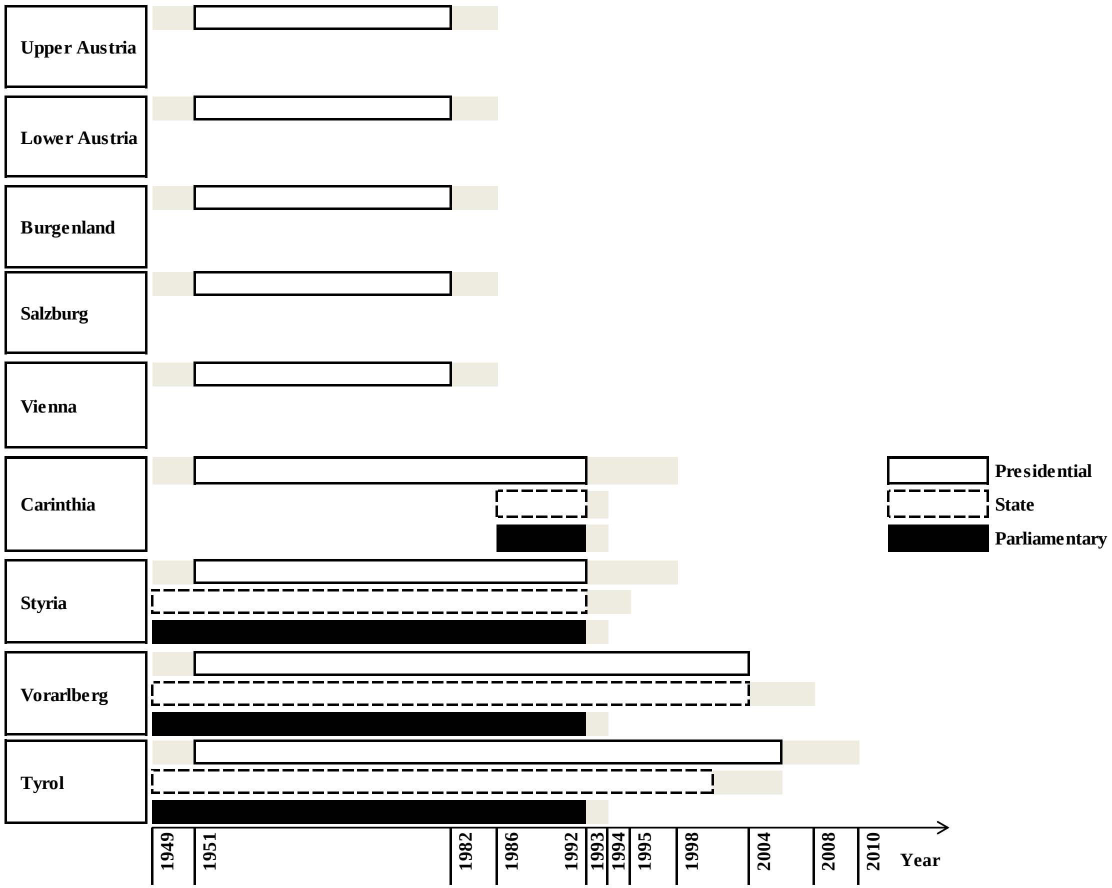
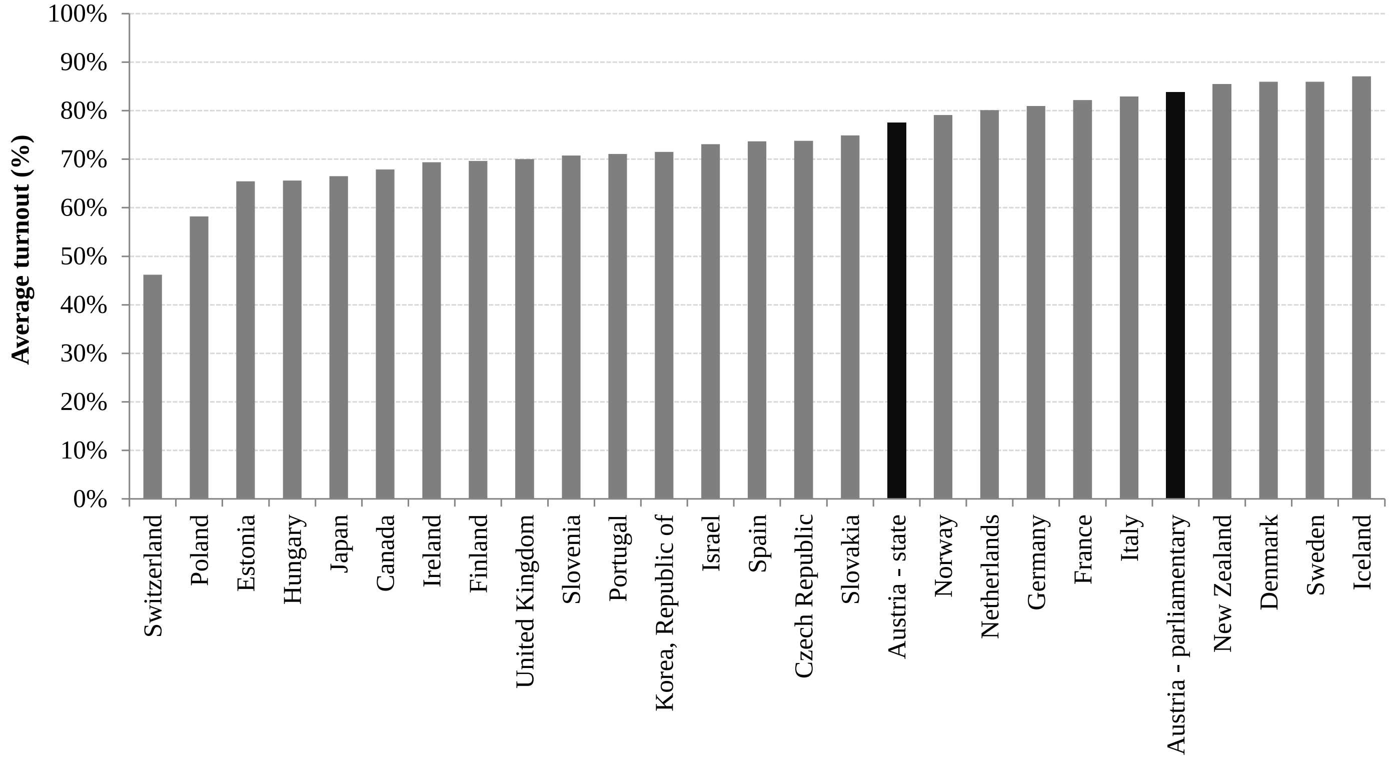


Figure 2.2: Elections Under Compulsory and Voluntary Voting (1949-2010)



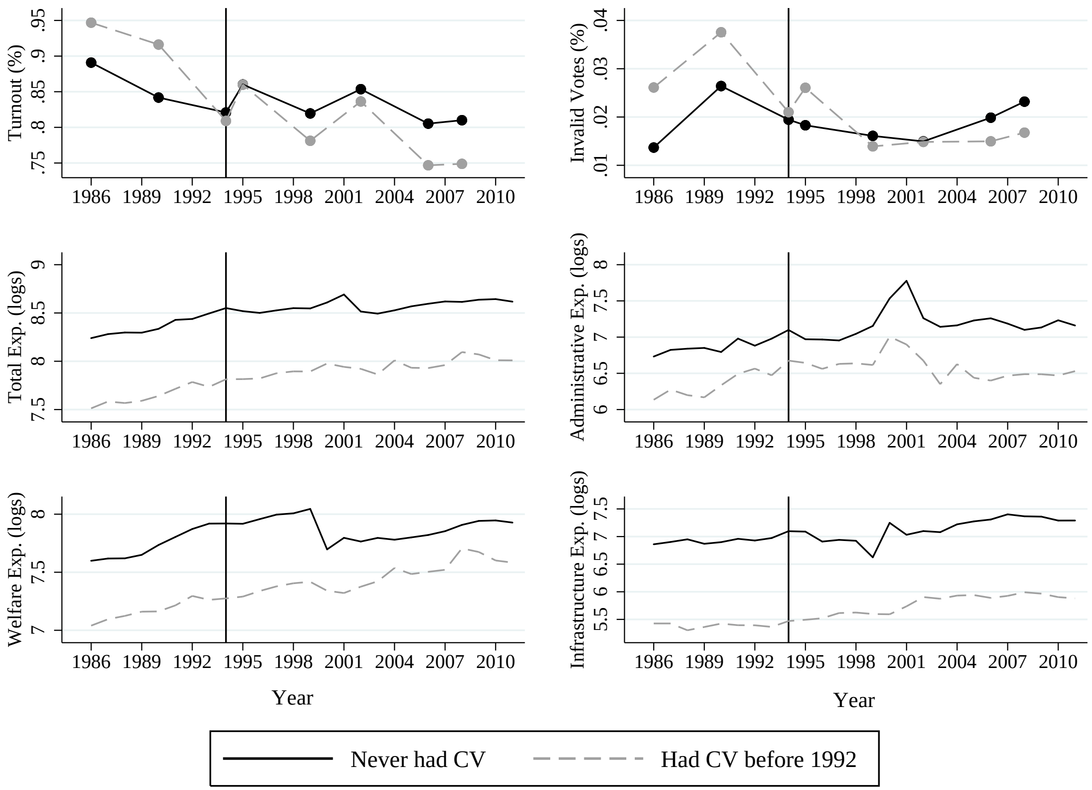
Notes: The legend marks the period in which elections with CV were held for each type of election. Shaded bars indicate that the enactment/abolition of the CV law was already in place, although no elections were held in that period.

Figure 2.3: Average Turnout in OECD Countries With Voluntary Voting (1979-2010)



Notes: Average turnout (as a % of registered voters) for all elections held in 1979-2010 in OECD countries with voluntary voting. We consider presidential elections in France and South Korea, and parliamentary elections in all others. In the case of Austria, we display the average turnout in parliamentary and state elections considering only elections in which voting was voluntary. Voter registration is compulsory and/or automatic (broadly speaking, e.g., it may be opt out) in all of these countries, making turnout as a % of registered voters comparable to turnout as a % of the voting age population (VAP). Details about each of the countries' voter registration procedures are provided by the ACE Electoral Knowledge Network at http://aceproject.org/epic-en/research/VR/spreadsheet_country. We have excluded the US, where registration is not automatic or mandatory. In US presidential elections, for example, turnout as a % of registered voters in this period was 85%, but only 56% when defined as a % of the VAP. Data on turnout in OECD countries was obtained from the Institute for Democracy and Electoral Assistance.

Figure 2.4: Turnout, Invalid Votes and Expenditures (1986-2012)



Notes: Turnout measures the percentage of registered voters who issued a vote, and invalid votes is the proportion of votes considered invalid. All state expenditures are expressed in millions of 2010 euros, in logs. The figures for turnout and invalid votes cover the parliamentary elections held in 1986-2008, and the figures for expenditures cover all the years in 1986-2012.

Table 2.1: Summary Statistics (1949-2012)

	Obs	Mean	Std. Dev	Min	Max
Panel A: Election Data (1949-2010)					
Parliamentary Elections					
Turnout (%)	171	0.90	0.07	0.70	0.98
Invalid Votes(%)	171	0.02	0.01	0.01	0.05
Votes Right (%)	171	0.52	0.09	0.22	0.75
Votes Left (%)	171	0.40	0.10	0.15	0.62
Votes Minor Parties (%)	171	0.09	0.09	0.00	0.49
Number of Parties	171	6.96	2.94	4.00	13.00
Vote Share Winner (%)	171	0.47	0.08	0.29	0.65
Margin of Victory (%)	171	0.12	0.09	0.00	0.40
State Elections					
Turnout (%)	121	0.86	0.09	0.61	0.98
Invalid Votes(%)	121	0.02	0.01	0.01	0.08
Votes Right (%)	121	0.53	0.11	0.21	0.76
Votes Left (%)	121	0.39	0.11	0.10	0.62
Votes Minor Parties (%)	121	0.07	0.08	0.00	0.50
Number of Parties	121	5.98	1.05	4.00	8.00
Vote Share Winner (%)	121	0.49	0.06	0.36	0.65
Margin of Victory (%)	121	0.10	0.07	0.00	0.34
Presidential Elections					
Turnout (%)	132	0.88	0.13	0.38	1.00
Invalid Votes(%)	135	0.04	0.02	0.01	0.11

Continued on next page

Table 2.1 – *Continued from previous page*

	Obs	Mean	Std. Dev	Min	Max
Panel B: Yearly State Data (1980-2012)					
Unemployment Rate (%)	297	0.06	0.02	0.00	0.10
Population (in thousands)	297	890	492	269	1717
Administrative Expenditures	**297**	**898.67**	**779.60**	**118.98**	**4303.17**
Representatives and gen. admin	297	423.67	431.13	74.81	2280.97
Finance	297	475.00	485.26	42.20	3699.91
Welfare Expenditures	**297**	**1977.86**	**1415.89**	**341.19**	**6916.49**
Education, sports and science	297	636.42	399.04	138.80	1774.97
Social welfare and housing	297	701.01	521.13	106.37	2315.46
Health	297	569.51	548.88	81.82	2977.10
Arts, culture and religion	297	70.92	62.89	9.91	288.36
Infrastructure Expenditures	**297**	**763.11**	**1141.22**	**72.16**	**4818.05**
Roads and transport	297	230.64	211.70	47.55	1010.88
Public order and security	297	26.03	39.55	1.21	163.66
Promotion of the economy	297	121.33	64.87	21.13	318.33
Services	297	385.11	933.29	0.98	4055.52

Notes: State-level election data covers all elections held from 1949 to 2010. *Turnout measures* the percentage of registered voters who issued a vote, and *invalid votes* is the proportion of ballots considered invalid. Vote shares for the *right* and *left* are the percentage of valid votes that went to ÖVP + FPÖ and SPÖ + KPÖ, respectively. Votes shares for *minor parties* are the percentage of valid votes received by other smaller parties. The *vote share of the winner* is the percentage of valid votes obtained by the highest ranking party in each state, and *margin of victory* is the difference in vote shares between the highest ranking party and the runner-up. Expenditure, unemployment, and population data at the state-level cover all the years from 1980 to 2012. All state expenditures are expressed in millions of 2010 euros.

Table 2.2: Descriptive Statistics: 1986 and 2003 Austrian Social Survey

	Obs	Mean	Std. Dev	Min	Max
Turnout					
Voted in Last Parliamentary Elections (%)	3726	0.87	0.34	0.00	1.00
Political Party of Preference					
Left (%)	3670	0.29	0.45	0.00	1.00
Right (%)	3670	0.34	0.47	0.00	1.00
Minor Parties (%)	3670	0.07	0.25	0.00	1.00
No Party Preference (%)	3670	0.30	0.46	0.00	1.00
Not member of a Political Party (%)	3693	0.82	0.39	0.00	1.00
Interest in Politics and Information					
Uninterested in Politics (%)	3723	0.36	0.48	0.00	1.00
Mildly Interested in Politics(%)	3723	0.40	0.49	0.00	1.00
Very Interested in Politics (%)	3723	0.24	0.43	0.00	1.00
Does not read newspaper regularly	3705	0.31	0.46	0.00	1.00
Socioeconomic Variables					
Age	3726	46.04	16.71	18.00	92.00
Female	3726	0.59	0.49	0.00	1.00
Household Income (in 2003 Euros)	2918	1797.95	963.54	180.00	4341.90
Number of members in household	3726	2.85	1.57	1.00	9.00
Employed (%)	3726	0.49	0.50	0.00	1.00
Unemployed (%)	3726	0.03	0.17	0.00	1.00
Retired (%)	3726	0.27	0.44	0.00	1.00
Educational Attainment					
Compulsory Schooling (%)	3726	0.65	0.48	0.00	1.00
Vocational Middle School (%)	3726	0.13	0.34	0.00	1.00
High School (%)	3726	0.16	0.37	0.00	1.00
College (%)	3726	0.06	0.24	0.00	1.00

Notes: The sample includes all individuals in the 1986 and 2003 Austrian Social Survey who reported whether they voted in the last parliamentary elections (1983 and 2002) and were of voting age. *Political party of preference* specifies the party the respondent identifies with (left if SPÖ or KPÖ, right if ÖVP or FPÖ, no party preference if the individual does not identify with any party, and minor parties otherwise), and not a member of a political party is a dummy for whether the individual has no party affiliation. Individuals are separated into three categories according to whether they manifest to be uninterested, mildly or very interested in politics. The Austrian Social Survey separates household income into 21 different categories. To make the figures comparable across periods, we imputed household income as the midpoint of the category into which individuals fell, and converted the 1986 mid-point into 2003 euros. Educational variables are mutually exclusive dummies for the maximum educational attainment. The educational category of "Vocational Middle School" corresponds with "berufsbildende mittlere Schule" (or "bms") in the data.

Table 2.3: Effect of CV on Turnout and Invalid Votes (1949-2010)

	Parliamentary	State	Presidential	Pooled
Panel A: Turnout (%) as Dependent Variable				
CV	0.065**	0.172**	0.095***	
	(0.020)	(0.059)	(0.022)	
	[0.002]	[0.000]	[0.012]	
CV * Parliamentary				0.066**
				(0.021)
				[0.006]
CV * State				0.081**
				(0.024)
				[0.000]
CV * Presidential				0.091***
				(0.021)
				[0.006]
Observations	171	121	132	424
R^2	0.956	0.953	0.961	0.934
Mean Turnout (if CV=0)	0.879	0.833	0.763	0.844
Panel B: Invalid Votes (%) as Dependent Variable				
CV	0.005	0.019*	0.019**	
	(0.004)	(0.008)	(0.007)	
	[0.200]	[0.000]	[0.000]	
CV * Parliamentary				0.009***
				(0.002)
				[0.018]
CV * State				0.013***
				(0.003)
				[0.054]
CV * Presidential				0.018***
				(0.005)
				[0.000]
Observations	171	121	135	427
R^2	0.787	0.834	0.835	0.831
Mean Invalid Votes (if CV=0)	0.015	0.018	0.038	0.020

Notes: Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects, and state-specific linear trends (at the election level). An observation in these regressions is a state-election. The sample in columns (1)-(3) includes all parliamentary, state, and presidential elections from 1949-2010, respectively. The pooled regressions in column (4) include all three types of elections from 1949-2010, with their corresponding election type dummies. In column 4, we include three different election-numbered state linear trends. *Turnout* measures the percentage of registered voters who issued a vote, and *invalid votes* is the proportion of ballots considered invalid. *CV* is a dummy for whether voting was mandatory in the state and election. For three of the presidential elections, turnout in the data is listed at over 100%, so they are excluded. Results are qualitatively robust to including these points. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.4: Effect of CV on Expenditures (1980-2012)

	Log-Levels				Log-Per Capita				% of Total		
	Total	Admin.	Welfare	Infrast.	Total	Admin.	Welfare	Infrast.	Admin.	Welfare	Infrast.
Panel A: Parliamentary Elections											
CV	0.031	0.006	0.035	0.066	0.008	-0.018	0.011	0.043	-0.008	0.001	0.007
	(0.017)	(0.112)	(0.034)	(0.097)	(0.037)	(0.121)	(0.048)	(0.088)	(0.027)	(0.025)	(0.015)
	[0.124]	[0.893]	[0.284]	[0.464]	[0.883]	[0.909]	[0.737]	[0.609]	[0.757]	[0.981]	[0.659]
R^2	0.991	0.949	0.990	0.979	0.930	0.792	0.895	0.942	0.590	0.791	0.875
Panel B: State Elections											
CV	-0.019	-0.168	0.033	-0.029	0.023	-0.126	0.075	0.013	-0.040*	0.034	0.006
	(0.037)	(0.104)	(0.036)	(0.101)	(0.058)	(0.116)	(0.059)	(0.088)	(0.021)	(0.022)	(0.017)
	[0.633]	[0.096]	[0.324]	[0.775]	[0.649]	[0.396]	[0.202]	[0.825]	[0.040]	[0.130]	[0.701]
R^2	0.991	0.950	0.990	0.979	0.930	0.795	0.897	0.942	0.601	0.796	0.875
Observations	297	297	297	297	297	297	297	297	297	297	297
Mean Dep. Variable	7.878	6.449	7.342	5.937	-5.649	-7.078	-6.184	-7.590	0.248	0.592	0.160
P-Value (equality across election types)	0.316	0.304	0.956	0.581	0.870	0.585	0.478	0.836	0.360	0.260	0.965

Notes: Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects, and state-specific linear trends (at the year level). An observation in these regressions is a state-year, and the sample is composed of all the years from 1980-2012. The dependent variable is yearly real spending by the state government, and the independent variable in Panels A and B is whether voting in the corresponding electoral period was compulsory for parliamentary or state elections, respectively. Spending totals and spending for each subcategory (administrative, welfare, and infrastructure) is in logs in columns (1)-(4) and in per capita logs in columns (5)-(8). Yearly spending for each subcategory is expressed as a percentage of total state spending in columns (9)-(11). * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.5: Effect of CV on Electoral Outcomes (1949-2010)

	Left (%)	Right (%)	Minor Parties (%)	Num. of Parties	Share Winner (%)	Margin of Victory
Panel A: Parliamentary Elections						
CV	-0.014	0.018	-0.003	0.413	0.020	0.040
	(0.009)	(0.017)	(0.018)	(0.336)	(0.017)	(0.026)
	[0.130]	[0.468]	[0.845]	[0.312]	[0.296]	[0.118]
Observations	171	171	171	171	171	171
R^2	0.984	0.922	0.902	0.965	0.886	0.765
Mean Dep. Variable	0.400	0.515	0.085	6.959	0.465	0.120
Panel B: State Elections						
CV	0.035	0.001	-0.036	-0.052	0.014	0.004
	(0.046)	(0.038)	(0.043)	(0.514)	(0.036)	(0.044)
	[0.623]	[0.977]	[0.352]	[0.915]	[0.643]	[0.891]
Observations	121	121	121	121	121	121
R^2	0.967	0.915	0.848	0.823	0.870	0.790
Mean Dep. Variable	0.393	0.535	0.072	5.975	0.494	0.101
P-Value (equality across election types)	0.210	0.720	0.491	0.368	0.881	0.551

Notes: Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects, and state-specific linear trends (at the election level). In column 4, we include three different election-numbered state linear trends. An observation in these regressions is a state-election. The sample in Panels A and B includes all parliamentary and state elections from 1949-2010, respectively. Vote shares for the *left* and *right* are the percentage of valid votes that went to SPÖ + KPÖ and ÖVP + FPÖ, respectively. Votes shares for *minor parties* are the percentage of valid votes received by other smaller parties. *Number of parties* is the number of parties participating in each election and state. The *vote share of the winner* is the percentage of valid votes obtained by the highest-ranking party in each state, and *margin of victory* is the difference in vote shares between the highest-ranking party and the runner-up. *CV* is a dummy for whether voting was mandatory in the state in that particular election. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.6: Individual-level Impact of CV on Turnout: Heterogeneity by Voter Characteristics

Dependent Variable: Voted in Last Parliamentary Elections			
1) Effect of CV on Turnout		**5) Party Membership**	
CV	0.055* (0.028) [0.209]	Not Party Member * CV	0.066* (0.029) [0.235]
		Party Member * CV	0.011 (0.028) [0.768]
2) Gender		**6) Informed. vs Uninformed**	
Female * CV	0.075* (0.035) [0.217]	Uninformed * CV	0.087** (0.033) [0.311]
Male * CV	0.027 (0.019) [0.255]	Informed * CV	0.043 (0.027) [0.225]
3) Education		**7) Interest in Politics**	
Compulsory Schooling * CV	0.049 (0.028) [0.215]	Uninterested * CV	0.082* (0.039) [0.253]
Vocational Middle School * CV	0.104** (0.044) [0.271]	Mildly Interested * CV	0.045 (0.024) [0.161]
High School or College * CV	0.038 (0.039) [0.343]	Very Interested * CV	0.017 (0.029) [0.576]

Continued on next page

Table 2.6 – *Continued from previous page*

Dependent Variable: Voted in Last Parliamentary Elections			
4) Income Quartile		**8) Political Preference**	
Income Q1 * CV	0.072*	Left * CV	0.039
	(0.034)		(0.030)
	[0.217]		[0.271]
Income Q2 * CV	0.075*	Right * CV	0.034
	(0.039)		(0.022)
	[0.271]		[0.245]
Income Q3 * CV	0.029	Minor Parties * CV	-0.017
	(0.034)		(0.153)
	[0.467]		[0.928]
Income Q4 * CV	0.037	No Party Preference * CV	0.148**
	(0.030)		(0.063)
	[0.321]		[0.077]

Notes: This table presents 8 separate regressions numbered 1-8. Except for regression 1, the coefficients shown are interactions of *CV* with individual characteristics. In regression 6, we use whether someone regularly reads the newspaper as a proxy for whether they are informed. Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. The dependent variable in all regressions is a dummy for whether the individual voted in the previous parliamentary elections. *CV* is a dummy for whether voting was compulsory for that election in the individual's state of residence. All regressions include baseline controls for age, age squared, gender, educational attainment, parents' education, working status, household size, community size, self-reported political preference, party membership, being informed, interest in politics, state fixed effects, and survey year fixed effects. Regression 4 includes income quartile controls (excluded from the other regressions because they sometimes have missing data). The sample includes all individuals in the 1986 and 2003 Austrian Social Survey who reported whether they voted in the last parliamentary elections (1983 and 2002) and were of voting age. Regression 4 has 2,647 observations, whereas all other regressions have 3,369 observations. * significant at 10%; ** significant at 5%; *** significant at 1%

2.7 Appendix Tables

Table 2.A1: Description of Expenditure Subcategories and Groupings

Administrative Expenses
i) Elected representatives and general administration:
State parliament, state government, state government delegations, sub-state governments, special offices, committees, pensions, personnel expenses, and other tasks of the public administration.

ii) Finance:
Capital assets and unincorporated foundations, financial allocations and grants, liabilities, budgetary compensation, and handover and takeover of the annual results.

Welfare Expenses
i) Education, sports and science:
Secondary education, vocational education and teacher formation, preschool education, education promotion, extracurricular educational activities for the youth, sports and extracurricular physical education, adult education, and research and science.

ii) Social welfare and promotion of house construction:
General public welfare, youth welfare, emergency funds, social and family policies, and housing subsidies.

iii) Health:
Health services, environmental protection, rescue and warning services, health worker training, public hospitals, hospitals operated by other legal entities, and veterinary medicine.

iv) Arts, culture and religion:
Fine arts, music and performing arts, literature and language, museums, heritage preservation, radio, press, films, and church affairs.

Infrastructure Expenses
i) Road construction, hydraulic engineering and transport:
Road construction and maintenance, hydraulic construction, flood protection, road/rail/water traffic, aviation, and postal and telecommunication services.

ii) Promotion of the economy:
Improvement and promotion of agriculture and forestry, promotion of energy, tourism, trade, commerce, and industry.

iii) Services:
Public services (water supply, lighting, waste management, etc.), residential and commercial buildings, and utility companies.

iv) Public order and security:
Public order, security and special police, firefighting, disaster relief and national defense.

Notes: Detailed breakdown of each category was obtained from Appendix 2 of the 787/1996 Ministry of Finance regulation on budgeting and accounts (Voranschlags- und Rechnungsabschlußverordnung VRV)

Table 2.A2: Interest in Politics, Information Acquisition and Party Membership: Austria vs. Other OECD Countries

	OECD Countries (WVS 2005-09)	Germany and Scandinavia (WVS 2005-09)	Switzerland (WVS 2007)	Austria (ASS 2003)
Interest in Politics				
Very interested	12.10%	16.20%	26.60%	26.24%
Somewhat interested	39.10%	49.30%	45.30%	39.37%
Not very/not at all interested	48.20%	33.90%	27.80%	34.39%
Newspaper Reading				
Reads the newspaper regularly / Used newspaper last week	74.80%	89.80%	91.40%	68.73%
Membership of a political party				
Member of a party	13.40%	11.50%	16.00%	12.41%

Notes: Information from other OECD countries comes from the 2005-09 wave of the World Values Survey. Selected samples: Canada 2005, Finland 2005, France 2006, Germany 2006, Hungary 2009, Italy 2005, Japan 2005, Netherlands 2005, New Zealand 2004, Norway 2007, Poland 2005, Slovenia 2005, South Korea 2005, Spain 2007, Sweden 2006, Switzerland 2007, United Kingdom 2005, United States 2006. For Austria, statistics come from the 2003 ASS. For interest in politics, the ASS asks "*How interested are you in politics? (1) Very interested, (2) rather interested, (3) mildly interested, (4) not very interested, or (5) uninterested.*", whereas the WVS asks: "*How interested would you say you are in politics? (1) Very interested, (2) somewhat interested, (3) not very interested or (4) not at all interested*". We categorize (1)-(2) from the ASS and (1) from the WVS as *Very interested*, whereas (4)-(5) from the ASS and (4) in the WVS are categorized as *Not very/not at all interested*. Regarding newspaper reading, the ASS asks "*How regularly do you read the newspaper? Regularly or not regularly?*" The WVS asks "*People use different sources to learn what is going on in their country and the world. For each of the following sources [daily newspaper], please indicate whether you used it last week or did not use it last week to obtain information.*" On political party membership, the 1986 wave of the ASS asks "*Are you member of a political party? Yes or no,*" whereas the 2003 wave asks "*Are you a member or do you have an active role in a political party? Member without an active role (1), I have an active role (2), or I'm not a party member (3).*" We consider individuals answering (1) or (2) as being members of a political party. The WVS question is: "*I am going to read off a list of voluntary organizations. For each one [political party], could you tell me whether you are an active member, an inactive member or not a member of that type of organization?*" We consider active and inactive members of a political party.

Table 2.A3: Effect of CV on Turnout and Invalid Votes (1979-2010)

	Parliamentary	State	Presidential	Pooled
Panel A: Turnout (%) as Dependent Variable				
CV	0.045**	0.125	0.110***	
	(0.017)	(0.096)	(0.024)	
	[0.058]	[0.005]	[0.006]	
CV * Parliamentary				0.085**
				(0.028)
				[0.000]
CV * State				0.124***
				(0.037)
				[0.000]
CV * Presidential				0.108***
				(0.024)
				[0.012]
Observations	90	61	71	222
R^2	0.945	0.964	0.974	0.936
Mean Turnout (if CV=0)	0.838	0.775	0.763	0.800
Panel B: Invalid Votes (%) as Dependent Variable				
CV	0.001	0.012	0.022**	
	(0.004)	(0.012)	(0.008)	
	[0.795]	[0.069]	[0.004]	
CV * Parliamentary				0.007
				(0.006)
				[0.266]
CV * State				0.019*
				(0.009)
				[0.024]
CV * Presidential				0.021**
				(0.008)
				[0.008]
Observations	90	61	72	223
R^2	0.779	0.911	0.877	0.863
Mean Invalid Votes (if CV=0)	0.017	0.020	0.038	0.023

Notes: Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects and state-specific linear trends (at the election level). An observation in these regressions is a state-election. The sample in columns (1)-(3) includes all parliamentary, state, and presidential elections from 1979-2010, respectively. The pooled regressions in column (4) include the three types of elections elections from 1979-2010, with their corresponding election type dummies. *Turnout* measures the percentage of registered voters who issued a vote, and *invalid votes* is the proportion of ballots considered invalid. *CV* is a dummy for whether voting was mandatory in the state and election. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.A4: Effect of Presidential CV on Expenditures (1980-2012)

	Log-Levels				Log-Per Capita				% of Total		
	Total	Admin.	Welfare	Infrast.	Total	Admin.	Welfare	Infrast.	Admin.	Welfare	Infrast.
CV	-0.016	-0.031	-0.020	-0.086	0.023	0.007	0.018	-0.048	-0.001	0.001	-0.000
	(0.037)	(0.084)	(0.050)	(0.097)	(0.051)	(0.091)	(0.065)	(0.095)	(0.019)	(0.021)	(0.017)
	[0.697]	[0.693]	[0.775]	[0.392]	[0.681]	[0.961]	[0.749]	[0.621]	[0.891]	[0.941]	[0.975]
R^2	0.991	0.949	0.990	0.979	0.930	0.792	0.895	0.942	0.590	0.791	0.875
Observations	297	297	297	297	297	297	297	297	297	297	297
Mean Dep. Variable	7.878	6.449	7.342	5.937	-5.649	-7.078	-6.184	-7.590	0.248	0.592	0.160

Notes: This table is similar to Table 2.4 in the main text except the regressor of interest is whether there is CV in presidential elections (instead of parliamentary or state elections). See notes of Table 2.4 for details. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.A5: Effect of CV on All Expenditure Categories – Parliamentary Elections (1980-2012)

	Administrative		Welfare				Infrastructure			
	General	Finance	Education	Social	Health	Arts	Transport	Security	Economy	Services
Panel A: Log-Levels										
CV	0.036	-0.009	0.026	-0.036	0.138	-0.031	-0.031	-0.080	0.207	-0.024
	(0.035)	(0.192)	(0.020)	(0.072)	(0.077)	(0.053)	(0.112)	(0.147)	(0.115)	(0.481)
	[0.376]	[1.000]	[0.180]	[0.717]	[0.302]	[0.513]	[0.801]	[0.633]	[0.084]	[0.949]
R^2	0.993	0.887	0.999	0.988	0.925	0.991	0.935	0.978	0.917	0.940
Panel B: Log-Per Capita										
CV	0.013	-0.033	0.002	-0.059	0.114	-0.055	-0.055	-0.104	0.183	-0.048
	(0.045)	(0.200)	(0.034)	(0.070)	(0.096)	(0.064)	(0.109)	(0.140)	(0.109)	(0.476)
	[0.801]	[0.903]	[0.999]	[0.551]	[0.464]	[0.368]	[0.675]	[0.549]	[0.122]	[0.927]
R^2	0.940	0.698	0.920	0.897	0.818	0.952	0.728	0.945	0.860	0.915
Panel C: % of Total										
CV	0.000	-0.008	-0.003	-0.013	0.019	-0.001	0.000	-0.000	0.012	-0.005
	(0.005)	(0.029)	(0.005)	(0.017)	(0.013)	(0.001)	(0.006)	(0.001)	(0.010)	(0.007)
	[1.000]	[0.757]	[0.557]	[0.663]	[0.356]	[0.258]	[0.975]	[0.737]	[0.178]	[0.298]
R^2	0.797	0.521	0.878	0.715	0.728	0.876	0.762	0.917	0.760	0.935
Mean Dep. Variable	0.119	0.129	0.199	0.206	0.167	0.020	0.067	0.006	0.045	0.042
Observations	297	297	297	297	297	297	297	297	297	297

Notes: Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects, and state-specific linear trends (at the year level). An observation in these regressions is a state-year, and the sample is composed of all the years in 1980-2012. The independent variable in all elections is whether voting in the corresponding electoral period was compulsory for parliamentary elections The dependent variable is the log of yearly real spending by the state government in Panel A, the log of yearly per capita spending in Panel B, and spending in that particular category as a % of total spending by the state government in Panel C. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.A6: Effect of CV on All Expenditure Categories – State Elections (1980-2012)

	Administrative		Welfare				Infrastructure			
	General	Finance	Education	Social	Health	Arts	Transport	Security	Economy	Services
Panel A: Log-Levels										
CV	0.012	-0.298	-0.002	0.015	0.107	0.029	-0.059	-0.011	-0.025	-0.042
	(0.035)	(0.189)	(0.016)	(0.042)	(0.104)	(0.043)	(0.092)	(0.091)	(0.097)	(0.210)
	[0.777]	[0.122]	[0.943]	[0.729]	[0.370]	[0.442]	[0.505]	[0.921]	[0.811]	[0.833]
R^2	0.993	0.890	0.999	0.988	0.924	0.991	0.935	0.978	0.914	0.940
Panel B: Log-Per Capita										
CV	0.054	-0.256	0.041	0.057	0.149	0.072	-0.017	0.031	0.017	0.001
	(0.049)	(0.198)	(0.042)	(0.065)	(0.113)	(0.066)	(0.073)	(0.092)	(0.079)	(0.238)
	[0.328]	[0.234]	[0.398]	[0.306]	[0.248]	[0.316]	[0.805]	[0.785]	[0.781]	[0.961]
R^2	0.941	0.704	0.921	0.898	0.819	0.952	0.728	0.944	0.857	0.915
Panel C: % of Total										
CV	0.004	-0.044*	0.003	0.007	0.023	0.001	-0.007	-0.000	-0.002	0.014
	(0.006)	(0.022)	(0.007)	(0.014)	(0.013)	(0.001)	(0.007)	(0.001)	(0.007)	(0.017)
	[0.613]	[0.040]	[0.655]	[0.657]	[0.174]	[0.412]	[0.406]	[0.851]	[0.813]	[0.356]
R^2	0.798	0.534	0.878	0.713	0.731	0.877	0.765	0.917	0.754	0.936
Mean Dep. Variable	0.119	0.129	0.199	0.206	0.167	0.020	0.067	0.006	0.045	0.042
Observations	297	297	297	297	297	297	297	297	297	297

Notes: Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects, and state-specific linear trends (at the year level). An observation in these regressions is a state-year, and the sample is composed of all the years from 1980-2012. The independent variable in all elections is whether voting in the corresponding electoral period was compulsory for state elections. The dependent variable is the log of yearly real spending by the state government in Panel A, the log of yearly per capita spending in Panel B, and spending in that particular category as a % of total spending by the state government in Panel C. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.A7: Instrumental Variables Regression: Effect of Turnout on Expenditures (1980-2012)

	Log-Levels				Log-Per Capita				% of Total		
	Total	Admin.	Welfare	Infrast.	Total	Admin.	Welfare	Infrast.	Admin.	Welfare	Infrast.
Panel A: Parliamentary Elections											
Turnout (%)	0.556	0.099	0.618	1.180	0.138	-0.318	0.201	0.763	-0.151	0.022	0.129
	(0.384)	(1.734)	(0.500)	(1.447)	(0.603)	(1.796)	(0.748)	(1.360)	(0.394)	(0.381)	(0.229)
	[0.124]	[0.893]	[0.284]	[0.464]	[0.883]	[0.909]	[0.737]	[0.609]	[0.757]	[0.981]	[0.659]
First Stage F-Statistic	10.521	10.521	10.521	10.521	10.521	10.521	10.521	10.521	10.521	10.521	10.521
R^2	0.991	0.949	0.989	0.978	0.930	0.793	0.894	0.941	0.590	0.791	0.875
Panel B: State Elections											
Turnout (%)	-0.141	-1.235*	0.238	-0.213	0.170	-0.923	0.550	0.098	-0.296*	0.253*	0.043
	(0.224)	(0.698)	(0.251)	(0.615)	(0.355)	(0.801)	(0.358)	(0.561)	(0.156)	(0.148)	(0.109)
	[0.633]	[0.096]	[0.324]	[0.775]	[0.649]	[0.396]	[0.202]	[0.825]	[0.040]	[0.130]	[0.701]
First Stage F-Statistic	10.549	10.549	10.549	10.549	10.549	10.549	10.549	10.549	10.549	10.549	10.549
R^2	0.991	0.947	0.990	0.979	0.930	0.786	0.894	0.942	0.571	0.783	0.875
Observations	297	297	297	297	297	297	297	297	297	297	297
Mean Dep. Variable	7.878	6.449	7.342	5.937	-5.649	-7.078	-6.184	-7.590	0.248	0.592	0.160

Notes: Standard errors clustered by state in parentheses. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects, and state-specific linear trends (at the year level). An observation in these regressions is a state-year, and the sample is composed of all the years from 1980-2012. The dependent variable is yearly real spending by the state government, and the independent variable in Panels A, B, and C is turnout in the parliamentary, state, and presidential elections for the corresponding electoral period, respectively. Spending totals and spending for each subcategory (administrative, welfare, and infrastructure) is in logs in columns (1)-(4) and in per capita logs in columns (5)-(8). Yearly spending for each subcategory is expressed as a percentage of total state spending in columns (9)-(11). All equations are estimated using an instrumental variables regression. *Turnout* measures the percentage of registered voters who issued a vote, and is instrumented with a dummy for whether voting was mandatory in the state and electoral period. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.A8: Effect of the 1992 Elimination of CV on Turnout, Invalid Votes, and Spending

	Turnout (%)	Invalid Votes (%)	Total Exp. (in logs)	Admin. Exp. (in logs)	Welfare Exp. (in logs)	Infrast. Exp. (in logs)
CV	0.098***	0.013**	-0.006	0.059	-0.076	0.110
	(0.014)	(0.004)	(0.066)	(0.050)	(0.058)	(0.187)
	[0.008]	[0.020]	[0.946]	[0.245]	[0.263]	[0.564]
Observations	72	72	234	234	234	234
R^2	0.921	0.723	0.983	0.936	0.982	0.954
Mean Dep. Variable	0.835	0.020	7.970	6.546	7.434	6.010

Notes: Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. All regressions control for state population and unemployment rate, and include state fixed effects, and year fixed effects. An observation is a state-election in columns (1)-(2) and a state-year in columns (3)-(6). The sample includes all parliamentary elections held in 1986-2008 in columns (1)-(2), and yearly state-level spending for 1986-2012 in columns (3)-(6). *Turnout* measures the percentage of registered voters who issued a vote, and i*invalid votes* is the proportion of ballots considered invalid. Real spending in total and for each subcategory (administrative, welfare, and infrastructure) is in logs. *CV* is a dummy for whether voting was compulsory in the corresponding parliamentary elections in columns (1)-(2), and a dummy for whether voting in parliamentary elections was compulsory in the corresponding electoral period in columns (3)-(6). * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.A9: Robustness Check: Effect of CV on Turnout and Invalid Votes

	Turnout (%)	Invalid Votes (%)
CV (t+1) * Parliamentary	-0.017	-0.000
	(0.020)	(0.005)
	[0.348]	[0.919]
CV (t) * Parliamentary	0.098**	0.009
	(0.037)	(0.007)
	[0.018]	[0.246]
CV (t-1) * Parliamentary	-0.011	0.003
	(0.020)	(0.003)
	[0.573]	[0.168]
CV (t+1) * State	-0.003	0.001
	(0.014)	(0.008)
	[0.869]	[0.895]
CV (t) * State	0.109*	0.018*
	(0.049)	(0.008)
	[0.000]	[0.056]
CV (t-1) * State	-0.017	-0.007
	(0.055)	(0.004)
	[0.761]	[0.270]
CV (t+1) * Presidential	-0.026*	-0.003
	(0.012)	(0.006)
	[0.034]	[0.613]
CV (t) * Presidential	0.078***	0.012*
	(0.017)	(0.006)
	[0.006]	[0.032]
CV (t-1) * Presidential	-0.028	-0.001
	(0.017)	(0.004)
	[0.358]	[0.741]
Observations	388	391
R^2	0.935	0.853
Mean Dep. Variable (if CV=0)	0.844	0.020

Notes: Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects, state-specific linear trends (at the election level), and election type dummies. An observation in these regressions is a state-election, and the sample includes all parliamentary, state, and presidential elections from 1949-2010. *Turnout* measures the percentage of registered voters who issued a vote, and *invalid votes* is the proportion of ballots considered invalid. *CV* dummies and their lags and leads indicate whether voting was mandatory in a state and election. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.A10: Robustness Check: Effect of CV on Expenditures

	Log-Levels				Log-Per Capita				% of Total		
	Total	Admin.	Welfare	Infrast.	Total	Admin.	Welfare	Infrast.	Admin.	Welfare	Infrast.
Panel A: Parliamentary Elections											
CV (t+1)	0.017	-0.079	0.079	0.084	-0.021	-0.116	0.042	0.047	-0.031	0.031	0.001
	(0.047)	(0.143)	(0.082)	(0.140)	(0.072)	(0.149)	(0.105)	(0.133)	(0.038)	(0.040)	(0.026)
	[0.741]	[0.603]	[0.484]	[0.573]	[0.731]	[0.454]	[0.745]	[0.737]	[0.517]	[0.535]	[0.999]
CV (t)	-0.011	-0.023	-0.020	0.015	-0.008	-0.020	-0.017	0.018	-0.003	-0.005	0.008
	(0.018)	(0.117)	(0.022)	(0.059)	(0.022)	(0.121)	(0.018)	(0.055)	(0.027)	(0.022)	(0.007)
	[0.444]	[0.803]	[0.366]	[0.719]	[0.625]	[0.815]	[0.270]	[0.661]	[0.875]	[0.837]	[0.200]
CV (t-1)	0.064	0.084	0.059	0.051	0.036	0.055	0.030	0.022	0.004	-0.003	-0.001
	(0.035)	(0.171)	(0.038)	(0.074)	(0.027)	(0.153)	(0.053)	(0.076)	(0.038)	(0.036)	(0.014)
	[0.128]	[0.569]	[0.206]	[0.456]	[0.136]	[0.671]	[0.689]	[0.749]	[0.883]	[0.887]	[0.889]
R^2	0.991	0.949	0.990	0.979	0.930	0.795	0.895	0.942	0.595	0.794	0.875
Panel B: State Elections											
CV (t+1)	0.019	-0.071	0.055	-0.018	0.023	-0.067	0.060	-0.013	-0.024	0.029	-0.005
	(0.046)	(0.126)	(0.087)	(0.078)	(0.049)	(0.131)	(0.088)	(0.071)	(0.038)	(0.037)	(0.009)
	[0.695]	[0.739]	[0.763]	[0.787]	[0.671]	[0.753]	[0.675]	[0.831]	[0.915]	[0.893]	[0.585]
CV (t)	-0.020	-0.166	0.030	-0.029	0.022	-0.123	0.072	0.013	-0.039*	0.033*	0.006
	(0.039)	(0.092)	(0.034)	(0.101)	(0.057)	(0.106)	(0.054)	(0.088)	(0.018)	(0.018)	(0.017)
	[0.631]	[0.166]	[0.348]	[0.791]	[0.649]	[0.400]	[0.200]	[0.827]	[0.070]	[0.140]	[0.693]
CV (t-1)	0.001	-0.029	-0.007	-0.058	-0.001	-0.031	-0.008	-0.060	-0.005	0.000	0.005
	(0.061)	(0.146)	(0.095)	(0.130)	(0.074)	(0.166)	(0.095)	(0.129)	(0.030)	(0.036)	(0.025)
	[0.929]	[0.843]	[1.000]	[0.683]	[1.000]	[0.877]	[1.000]	[0.693]	[0.885]	[0.963]	[0.853]
R^2	0.991	0.950	0.990	0.979	0.930	0.796	0.898	0.942	0.604	0.799	0.875
Observations	297	297	297	297	297	297	297	297	297	297	297
Mean Dep. Variable	7.878	6.449	7.342	5.937	-5.649	-7.078	-6.184	-7.590	0.248	0.592	0.160

Notes: Standard errors clustered by state in parentheses, and cluster-robust wild-bootstrap p-values in square brackets. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects, and state-specific linear trends (at the year level). An observation in these regressions is a state-year, and the sample is composed of all the years from 1980-2012. The dependent variable is yearly real spending by the state government, and the independent variables in Panels A and B are whether voting in the corresponding, past, or future electoral period was compulsory for parliamentary or state elections, respectively. Spending totals and spending for each subcategory (administrative, welfare, and infrastructure) is in logs in columns (1)-(4) and in per capita logs in columns (5)-(8). Yearly spending for each subcategory is expressed as a percentage of total state spending in columns (9)-(11). * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.A11: Effect of CV on Electoral Outcomes (1979-2010)

	Left (%)	Right (%)	Minor Parties (%)	Num. of Parties	Share Winner (%)	Margin of Victory
Panel A: Parliamentary Elections						
CV	-0.020**	0.013	0.007	-0.091	0.005	0.034
	(0.007)	(0.025)	(0.027)	(0.522)	(0.016)	(0.031)
	[0.036]	[0.829]	[0.811]	[0.931]	[0.715]	[0.258]
Observations	90	90	90	90	90	90
R^2	0.993	0.900	0.925	0.982	0.886	0.712
Mean Dep. Variable	0.379	0.506	0.114	8.722	0.425	0.107
Panel B: State Elections						
CV	0.047	-0.017	-0.030	-0.538	-0.009	0.079
	(0.033)	(0.028)	(0.025)	(0.625)	(0.043)	(0.043)
	[0.047]	[0.053]	[0.003]	[0.003]	[0.812]	[0.267]
Observations	61	61	61	61	61	61
R^2	0.990	0.920	0.830	0.821	0.910	0.927
Mean Dep. Variable	0.363	0.549	0.088	6.607	0.478	0.106

Notes: This table is similar to Table 2.5 in the main text. The difference is the sample is restricted to 1979-2010. See notes of Table 2.5 for details. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.A12: Instrumental Variables Regression: Effect of Turnout on Electoral Outcomes (1949-2010)

	Left (%)	Right (%)	Minor Parties (%)	Num. of Parties	Share Winner (%)	Margin of Victory
Panel A: Parliamentary Elections						
Turnout (%)	-0.221	0.270	-0.049	6.333	0.302	0.609**
	(0.142)	(0.249)	(0.237)	(5.771)	(0.197)	(0.242)
	[0.130]	[0.468]	[0.845]	[0.312]	[0.296]	[0.118]
Observations	171	171	171	171	171	171
R^2	0.983	0.922	0.902	0.965	0.879	0.750
Mean Dep. Variable	0.400	0.515	0.085	6.959	0.465	0.120
Panel B: State Elections						
Turnout (%)	0.203	0.005	-0.208	-0.303	0.081	0.021
	(0.147)	(0.142)	(0.141)	(1.886)	(0.143)	(0.166)
	[0.623]	[0.977]	[0.352]	[0.915]	[0.643]	[0.891]
Observations	121	121	121	121	121	121
R^2	0.965	0.915	0.841	0.823	0.862	0.790
Mean Dep. Variable	0.393	0.535	0.072	5.975	0.494	0.101

Notes: Standard errors clustered by state in parentheses. All regressions control for state population and unemployment rate, and include state fixed effects, year fixed effects, and linear state election trends. An observation in these regressions is a state-election. The sample in Panels A and B includes all parliamentary elections and state elections from 1949-2010, respectively. Vote shares for the *left* and *right* are the percentage of valid votes that went to SPÖ + KPÖ and ÖVP + FPÖ, respectively. Votes shares for *minor parties* are the percentage of valid votes received by other smaller parties. *Number of parties* is the number of parties participating in each election and state. The *vote share of the winner* is the percentage of valid votes obtained by the highest ranking party in each state, and *margin of victory* is the difference in vote shares between the highest ranking party and the runner-up. All equations are estimated using an instrumental variables regression. *Turnout* measures the percentage of registered voters who issued a vote, and is instrumented with a dummy for whether voting was mandatory in the state and election. * significant at 10%; ** significant at 5%; *** significant at 1%

2.8 Further Background on CV in Austria

2.8.1 Compulsory Voting in Austria before 1945

The debate concerning the introduction of CV in Austria goes back to the enfranchisement of women in 1918. Conservative parties feared that their women supporters would not be as politically active and easy to mobilize as women who supported the Social Democrats, who had advocated for universal voting rights. CV was therefore seen as an instrument for conserving their power. Informal accounts mention that during the debates regarding the implementation of CV, conservatives put forward the argument that participation in political decisions and public life was not only a right but a duty of every citizen.[36] Social Democrats were against its implementation, and thus a compromise was reached, leaving the prerogative of instating mandatory voting to the states. In 1919, before the elections for the Constituent National Assembly, provisions for CV were made in Vorarlberg and Tyrol.[37] When the 1920 constitution was amended in December 1929, it became up to each state to decide whether voting was compulsory or not in national parliament and state parliament elections.[38]

The 1920 constitution, which was parliamentary in nature, underwent other important changes in 1929. The responsibilities of the president were broadened, and the election of the president became determined by popular vote rather than by decision of the members of the legislature. Furthermore, voting in presidential elections became mandatory in the whole country.[39] Although the first election was supposed to occur in 1931, due to the worldwide economic depression, political parties decided to suspend the elections and reelect the incumbent president. In May 1934, the Fascist ruling party repealed the 1929 constitution, but after World War II, in May 1945, the 1920 constitution (with its 1929 amendments) was reinstated. Thus, both the country-wide provisions for mandatory voting in

[36] http://www.onb.ac.at/ariadne/projekte/frauenwaehlet/Raum07.html, last accessed March 16, 2016

[37] http://www.parlament.gv.at/PERK/HIS/WAHL/REGEL/index.shtml, last accessed March 16, 2016.

[38] Federal Constitution of December 1929 (B-VG) Articles 26 (1) and 95 (1)-(2).

[39] Federal Constitution of December 1929 (B-VG) Article 60 (1).

presidential elections and the state-determined CV in national parliament and state parliament elections were restored. In spite of this, and probably due to the post-war chaos, the 1945 national and state parliament elections were carried out according to a national law made specifically for this election.[40] Thus, voting in the 1945 elections was optional for individuals in all states, including Tyrol and Vorarlberg. Only in the next election for national and state parliament, both held in 1949, did Vorarlberg and Tyrol re-implement CV. Furthermore, the state of Styria also enacted its own CV law for these elections.[41]

2.8.2 Fines for Abstention under Compulsory Voting

Maximum fines for abstention in presidential and parliamentary elections with CV were established by the National Parliament, whereas state parliaments had the authority for establishing maximum fines for non-voting in state elections. In all three election types, however, the actual fine amounts and their enforcement were left to local governments. As described in Section 2.2.2, abstention penalties were extremely rare, since the law allowed for a wide range of excuses for not voting.[42] Although there is no comprehensive information on the exact fines that individuals were charged with in the few cases in which these were enforced, anecdotal evidence suggests that fines were in fact much lower than the ceilings set by law.

Since each state had the authority for establishing maximum fines for non-voting in state elections, there was substantial variation across states (and time) in these maximum fines. When CV for state elections was established in Vorarlberg in 1919, fines varied depending on the socioeconomic status of the violator, ranging from 1 kronor(0.9 US dollars) to 50 Austrian kronor (44 USD).[43] The law establishing the value of these fines was modified over time, and in 1988, for

[40] Election Law 198 (Wahlgesetz) from October 1945.

[41] Styria Law 30 from July 11, 1949.

[42] Private correspondence between the authors and government officials confirmed that fines were enforced in only a handful of cases. Additionally, the website http://www.idea.int/vt/compulsory_voting.cfm describes the sanctions for not voting as being "weakly enforced" (accessed March 16, 2016).

[43] Vorarlberg State Law enacted in January 27, 1919, Article 2.

example, fines were capped at 10,000 schillings (1,413 USD).[44] Although maximum penalties were high, this ceiling was not binding, and in practice fines were significantly lower. Only in very few cases were non-voters effectively fined, with fines around 300-500 schillings (∼42-71 USD). Non-voters from Vorarlberg were asked by the mayor of their municipality to provide reasons for abstention, but were not required to provide any official proof. Those who did not comply with this request within a week and were reported to the authorities were granted an extra two weeks to provide a justification for their abstention. In 1949, punishment for abstention in state elections in Styria was set at a maximum of 1,000 schillings and four weeks of imprisonment, following the maximum sanctions for abstention in federal elections.[45] In the case of Tyrol, maximum fines for abstention in state parliament elections were always kept at 1,000 schillings, ranging due to inflation from around 506 USD in 1958 to 102 USD in 2002 when CV was eliminated (all in December 2015 values).[46] While the aforementioned states formally sanctioned abstention in state parliament elections, the enactment of CV in Carinthia in 1986 only explicitly set a punishment for abstention in federal elections (matching the corresponding federal laws). For state parliament elections, the law only states that abstainers must be sent a message from the government informing them about the importance of voting under a democratic state.[47]

Sanctions for abstention in presidential elections were initially capped at 1,000 schillings[48] (∼ 506 USD). In 2004, the last presidential election in which any state had CV, this sanction could amount to 72 euros (∼ 97 USD in December 2015).[49] The maximum fine for non-voting in parliamentary elections was also initially set at 1,000 schillings, but unlike presidential elections, the national law regulating parliamentary elections also established that failure to settle this fine was punish-

[44] Vorarlberg State Law number 60 enacted in December 14, 1988, Article 73(3). All figures in schillings are expressed in nominal terms. To express these in current dollars, the schilling values are updated to their 2015 value using the Austrian CPI, and then converted to dollars using the appropriate exchange rate.

[45] Styria State Law enacted in July 11, 1949, Article 1(3).

[46] Tyrol State Law number 27, enacted in July 29, 1949; Tyrol State Law number 20, enacted in July 5, 1965; and Tyrol State Law number 54, enacted in November 21, 1988.

[47] Carinthia state law issued in April 7, 1986, Article 3(3).

[48] 1957 Federal Presidential Election Law, Article 25.

[49] 2002 Federal Presidential Election Law, Article 23(3).

able by up to four weeks in jail.[50] In 1971, maximum sanctions for abstention in parliamentary elections were increased to 3,000 schillings (992 USD), but maximum imprisonment for not paying the fines was lowered to two weeks.[51]

2.8.3 Further Information on Elimination of CV Starting in 1994

Anecdotal information from the state legislature discussions on the elimination of CV shows that this repeal was triggered by the *de facto* null enforcement of the fines, and by the fact that parliamentary CV had already been repealed in 1992. Specific references can be found in Styria's state parliament session of January 26, 1993, and Tyrol's state parliamentary session of June 30, 2004.

2.9 Additional Discussion and Results

State spending. Beyond sharing of responsibilities, decision-making powers may also be shared across different levels of government for closely related areas OECD-06-ReformingFederalFiscal. In the description in Section 2.2.1 in the main text, we focus on the components of fiscal transfers. We use information from Tables 1 and 2 of lehner1997bundeslander to do our calculation about the share of transfers that are earmarked. It should be noted that federal transfers make up a majority but not all of state-level spending lehner1997bundeslander and there are also changes over time. Other sources focus on breakdowns for revenues instead of transfers. For example, OECD-06-ReformingFederalFiscal report that earmarked revenues are about one-third of total state revenues (net of additional fiscal arrangements between state and municipal governments) [p. 11]OECD-06-ReformingFederalFiscal. OECD1997 notes that "the proportion of their gross revenue which the [states] can spend at their own discretion rises to about 40 per cent" (this again focusing on revenues instead of transfers). Our point in the main text is that Austrian states still have autonomy over a considerable share of

[50] Federal Parliament Election Law, Article 105 (3).

[51] Article 109 (3) of the 1971 Federal Parliament Election Law.

the budget (in an absolute sense) and that our analysis would be able to pick up important changes in state-level spending were they to arise.[52]

Effect of the 1992 Elimination of CV on Turnout, Invalid Votes, and Spending. Table 2.A8 show the results of our difference-in-differences regression limiting our sample to the parliamentary elections in the electoral periods between 1986 and 2011, in which the only change in CV laws was federally enacted in 1992. This law forced Vorarlberg, Styria, Tyrol and Carinthia to eliminate CV in parliamentary elections.[53] The magnitude and statistical significance of the results is remarkably similar to those shown in Tables 2.3 and 2.4. The repeal of CV in 1992 causes a decrease in turnout in parliamentary elections of 9.8 percentage points, and an increase in invalid ballots of 1.3 percentage points. Likewise, in neither of our specifications do we find that CV affects fiscal policy. Due to the short time period covered in these regressions, we do not include state specific trends. Controlling for state-specific time trends, the repeal of CV in 1992 causes a decrease in turnout of 3.75 percentage points (significant at the 10% level), and an increase in the proportion of invalid ballots by 0.20 percentage points, although the latter is statistically insignificant. Furthermore, the results for our spending regressions are quantitatively similar when we control for state trends. These results suggest that any other changes in CV (besides the 1992 one) are unlikely to be correlated with trends in the main dependent variables.

Electoral Outcomes. As discussed in the main text, the SPÖ and ÖVP are the two major parties in Austria. We define "right-wing" parties as ÖVP and FPÖ and "left-wing" parties as SPÖ and KPÖ, although it is somewhat arbitrary to include

[52]Our point is not that Austrian states have a high level of fiscal autonomy compared to states in other federal countries (e.g., Austrian states lack significant tax powers OECD-06-ReformingFederalFiscal).

[53]The estimation equation is given by: $y_{st} = \alpha_0 + \alpha_1 CV_s * Pre_t + X_{st}\beta + \delta_s + \nu_t + \epsilon_{st}$. As in our previous specifications, y_{st} is an election outcome variable or expenditures in state s and year t; CV_s is a dummy variable indicating whether voting was compulsory in state s before the 1992 constitutional amendment, Pre_t is a dummy for the elections before 1992, X_{st} is a vector of state–year covariates (population and the unemployment rate), δ_s and ν_t are state and year fixed effects and ϵ_{st} is the error term. Our interest lies in the coefficient that measures the difference-in-differences between states with and without CV, before and after the reform, α_1. For comparison with previous tables, we introduce the "Pre" instead of "Post" dummy because after 1992, CV was repealed, rather than introduced.

the KPÖ in this list. The variables for the four parties are defined using the column headers marked "OEVP," "SPOE," "FPOE," and "KPOE" from the election data spreadsheets (which a research assistant created from the Ministry of Interior yearbooks). Thus, we ignore votes gained when smaller parties have run under different names and groupings over time. At the start of the sample period, there was the "Electoral Party of Independents," which was the predecessor to the FPÖ Manoschek02. We have defined this party as part of "minor parties," but our conclusions are robust to including it in the right-wing parties. In addition, we count the "linksblock" (left bloc) as part of minor parties. Our conclusions are also robust to defining the "left-wing" solely as the SPÖ (i.e., not counting the KPÖ as part of "left-wing" parties).

For our analysis of electoral outcomes, we do not study presidential elections because parties do not run as separate entities in those races. Instead, they form coalitions that cross party lines and change over time, making it impossible to identify the proportion of votes for right and left wing parties. In 1974, for example, the candidate nominated by the socialist SPÖ won the presidential election. This candidate was reelected in the 1980 elections, where he received support from the SPÖ but also from the right-wing ÖVP party.

With respect to our state and parliamentary election regressions, although we only report the results considering vote shares for left (SPÖ + KPÖ) and right wing (ÖVP + FPÖ) parties, we also run regressions using the individual vote shares of these parties and find no effect. We also checked whether there was any impact on voter polarization, and find that there is no effect of CV on the sum of vote shares for the two main parties (SPÖ and ÖVP).

CV Effects by Heterogeneity in Turnout. An additional test of our hypothesis is to explore the heterogeneity of the effect between low vs. high turnout states. If it is indeed the case that voters who turn out to vote only because of the introduction of CV do not have different preferences than those who vote despite the absence of CV, this effect should not depend on the baseline level of turnout, or even on the size of the "first stage" (i.e., the effect of CV on turnout). The results (available upon request) show that the effect of CV is larger in low turnout states, though the effect is not always statistically significant. This result is consistent with Funk2007, who documents that the effect of the abolition of CV (a law with low or no sanctions) in Swiss cantons is larger in places with lower baseline turnout. She argues that low baseline turnout is related to a social norm, thus in places in which the social norm was stronger, the effects of an expressive law were undermined. However, despite the heterogeneity that shows up in the first stage, the effect of CV on the composition and level of spending at the state level remains very close to zero and statistically insignificant in most cases.

Further Details on Elections. In Austria, elections occur for five bodies/offices: (1) The National Council (henceforth “parliamentary elections”), (2) State parliaments (“state elections”), (3) Federal President (“presidential elections”), (4) Municipal council, and (5) The European Parliament. Throughout the paper, we focus exclusively on the first three. In 2007, the voting age in Austria was lowered to 16 for nationwide elections, referenda, and plebiscites kritzinger2013austrian,wagner2012voting.

CV and Transmission of Political Information. One way that CV could affect turnout and other outcomes is by affecting political information. When voting is compulsory, this may change the incentive of voters to acquire political information and/or may also affect the incentive of parties to transmit information in political campaigns. To examine whether CV affects voters’ acquisition of information, we repeated the first regression in Table 6, but using newspaper reading as the outcome instead of turnout. We found no significant relation between CV and newspaper reading, though we recognize this is only a coarse measure of information acquisition. We do not have data on the campaign activities of political parties, so we cannot examine whether parties change their behavior. Overall, transmission of political information could be an important mechanism for our results, but it is difficult for us to examine this empirically.

Turnout disparity by income. Beyond the factors discussed in the Conclusion as reasons why the impacts of CV may vary by country, another factor is how large are the disparities in CV by income. See mahler2014electoral and http://www.politico.com/magazine/story/2015/01/income-gap-at-the-polls-113997 (last accessed in September 2016) for further evidence on disparities. Such disparities could affect how CV affects multiple outcomes, including turnout, spending, and electoral outcomes.

Chapter 3

Which Tail Matters? Inequality and Growth in Brazil

Joint with Stephan Litschig (National Graduate Institute for Policy Studies)

3.1 Introduction

A series of seminal theory papers propose different channels through which a society's degree of initial economic inequality might impact subsequent income per capita growth. These channels include aggregate savings and investment (Bourguignon, 1981), human and physical capital accumulation (Galor and Zeira, 1993; Banerjee and Newman, 1993; Aghion and Bolton, 1997; Barro, 2000; Galor and Moav, 2004), and income redistribution and social unrest (Persson and Tabellini, 1994; Alesina and Rodrik, 1994; Benabou, 1996; Esteban and Ray, 2000; Campante and Ferreira, 2007). Moreover, some of these mechanisms in theory have different implications for the effect of income inequality on growth, depending on whether the middle class is richer at the expense of the poor or the rich are richer at the expense of the middle class. Yet due to data limitations, evidence on mechanisms is extremely limited and existing cross-country evidence on the relationship between inequality and growth is largely inconclusive with effect estimates ranging from negative (Alesina and Rodrik, 1994; Persson and Tabellini, 1994; Perotti, 1996; Panizza, 2002) to zero (Voitchovsky, 2005; Ravallion, 2012) and positive (Forbes, 2000; Li et al., 1998).

Our paper investigates whether inequality originating from the lower as opposed to the upper tail of the income distribution has different effects on subsequent income per capita growth. Greater inequality as measured by commonly used metrics (e.g. the Gini coefficient) can result from higher dispersion in different parts of the income distribution, as illustrated in Figure 3.1. In Panel A, a theoretical redistribution of income from the bottom to the middle quintile (i.e., the transition from the Lorenz curve displayed in the solid line to that of the dashed line) implies higher overall income inequality as captured by the Gini coefficient. However, the exact same increase in overall inequality can be achieved by redistributing a portion of total income from the middle to the top quintile, as shown in Panel B. Most existing empirical work has effectively treated the variation in overall inequality the same, irrespective of the tail it originates from, despite the fact that the theory suggests that impacts on subsequent growth may differ. For example, in models with credit constraints and setup costs for human (Galor and Zeira, 1993) and physical capital investments (Barro, 2000), it is conceivable that only inequality in the lower tail matters for growth.

Consider a stylized economy with three groups of equal size (the poor, the middle class, and the rich) and the same income within each group. Now assume that the incomes of the poor and the middle class are initially too low to overcome the setup costs for investing in either human or physical capital. Put differently, both the poor and the middle class cannot borrow enough to make the relatively large investments that would be required to make a profit. Now consider another economy with the same income per capita but with higher inequality at the bottom, i.e. the middle class is richer while the poor are poorer. In this second economy, the middle class might be rich enough to overcome the setup costs and make profitable investments in human and physical capital, thus making the second economy richer than the first economy in the long run. Finally consider a third economy, again with identical income per capita but higher inequality at the top, i.e. the rich are richer at the expense of the middle class. Since human and physical capital investments are as constrained as in the more equal first economy, growth will be similarly limited.

Using sub-national data from Brazil over the 1970-2000 period we first establish that holding initial income per capita and a host of standard confounders constant, places with higher initial income inequality as measured by the Gini coefficient exhibit higher subsequent income per capita growth. Most of the effect materializes by 1991, i.e. there is only a level effect, not permanently higher growth. We then propose a simple approach to distinguish between the growth effects of inequality originating from the bottom versus the top of the initial income distribution. The key idea is to include quintile income shares instead of the Gini coefficient in an otherwise standard cross-sectional growth regression, allowing for hypothetical income redistributions from the two tails towards the (omitted) middle quintile while holding other income shares and mean income constant. We find that the positive association between overall initial inequality and subsequent growth is entirely driven by inequality in the lower tail of the income distribution: compared to more equal places, sub-national units with a 3 percentage point (one standard deviation) higher share of income going to the middle quintile at the expense of the bottom quintile experience about 3 percent higher income per capita by 2000. In contrast, places with a higher share of income going to the top quintile at the expense of the middle quintile get no growth boost at all compared to more equal places.

The differential effects of bottom versus top inequality are remarkably consistent with our evidence on human and physical capital accumulation. We find that places with a higher share of income held by the middle quintile at the expense of the bottom quintile experience higher subsequent growth in the real value of capital stocks held by businesses across all sectors of the economy. On the other hand, a higher share of income held by the top quintile at the expense of the middle quintile is not associated with increased physical capital accumulation. Moreover, human capital accumulation in places with higher inequality in the lower tail of the initial income distribution also outpaces places where the bottom quintile is richer, while inequality at the top of the distribution is uncorrelated with subsequent human capital growth. Other channels might also be at work. For example, higher inequality at the bottom might increase aggregate savings, and, in partly closed economies, aggregate investment. Higher inequality at the top on the other hand would affect aggregate savings only little if at all if the propensity to save is decreasing in income. Similarly, redistributive policies carried out at the local level may respond differently to lower- vs. upper-tail inequality. Unfortunately, we lack data on local savings or redistributive policies in the early 1970s.

Our paper builds on an extensive empirical literature linking overall income inequality and subsequent income per capita growth. Existing evidence is largely inconclusive and due to data limitations there is typically very little evidence on the mechanisms linking initial inequality to subsequent growth.[1] The most closely related study to ours is Voitchovsky (2005) which uses the 90/75 income percentile ratio as a proxy for inequality at the top of the income distribution, and the 50/10 income percentile ratio to proxy inequality at the bottom. For a sample of 21 developed countries, the study shows that under some specifications, inequality at the bottom is negatively correlated with growth, and inequality at the top has a positive correlation. The main conceptual difficulty with the Voitchovsky (2005) study is that the regression specifications typically include percentile ratios along with the Gini coefficient in the same equation. But a higher 90/75 income percentile ratio while keeping the Gini coefficient constant necessarily implies that inequality must be lower in other parts of the distribution. As a result, it is not clear what the coefficient on the 90/75 income percentile ratio is picking up. A similar issue arises in Ravallion (2012), which explores the impact of various parameters of the initial income distribution on income per capita growth and poverty reduction in a large sample of developing countries. The regression specification sometimes includes the initial poverty rate along with the Gini coefficient in the same equation. But holding initial income per capita constant, countries with a higher poverty rate are also those with higher overall inequality, as discussed in

[1]Early cross-country studies had found a negative relationship between inequality and growth (Alesina and Rodrik, 1994; Persson and Tabellini, 1994; Perotti, 1996). However, this negative correlation has proven to be non-robust to the inclusion of additional explanatory variables, and the lack of comparability of inequality measures across countries potentially results in attenuation and other measurement error bias (Forbes, 2000). This last issue has been overcome by a series of studies exploiting variation in inequality across sub-national units (e.g., Partridge, 1997; Panizza, 2002; Benjamin et al., 2011). Employing a higher quality dataset drawn from household surveys, Deininger and Squire (1998) find a negative cross-country correlation between initial income inequality and growth, whereas Barro (2000) finds a positive correlation between the Gini coefficient and average income growth in rich countries, but a negative correlation in poorer countries. A second strand of studies using higher quality data as well as panel data estimations find a positive (Partridge, 1997; Li et al., 1998; Forbes, 2000), negative (Panizza, 2002; Benjamin et al., 2011) and negative though often insignificant (Voitchovsky, 2005; Ferreira et al., 2014) correlation between Gini and growth.

that study. Moreover, holding both average income and overall inequality constant implies that countries with a higher poverty rate must have less inequality somewhere else in the distribution, which further complicates the interpretation of the coefficients.

The main contribution of our study is its conceptually straightforward approach to analyze the relationship between left and right tail inequality and subsequent outcomes. By replacing the Gini coefficient with the quintile income shares as our main regressors, we exploit variation in inequality originating from either tail while keeping initial average income and the other income shares constant. As illustrated in panels A and B of Figure 3.1, we exploit quantitatively identical differences in income inequality arising from opposing sides of the income distribution. This implies that the difference in effects of bottom versus top inequality we find is not driven by treating inequality in the two tails differently. And because our regressions hold income per capita constant, places with a lower share of income going to the poor and a higher share going to the middle class are places where the poor are poorer and the middle class is richer not only in relative but also in absolute terms. This is important because the credit market imperfections cum setup cost theory is based on absolute income levels. Another advantage of our setting is that we draw on homogeneously collected census data from a single country. Thus, unlike existing cross-country studies, we do not face a tradeoff between data quality and sample size, and our results are less prone to measurement error bias. Ours is also the first study to look at the effects of bottom versus top income inequality in a developing country context. An additional advantage of our setting is that by comparing sub-national units within the same country and state, we can abstract from differences in institutions at the federal and state level which might be correlated with initial inequality and income growth. Last but not least, our study also provides the first direct evidence on human and physical capital accumulation linking initial income distribution to subsequent economic growth.

A potential drawback compared to cross-country studies is that our results could be driven by migration, whereby places with high initial inequality at the bottom experience higher out-migration of the poor and thus higher income per capita among remaining residents in future periods for example. It turns out, however, that the effect of initial income inequality on in- or out-migration flows is close to negligible in practice as further discussed below. We also show that our results are unlikely to be driven by differential measurement error at the bottom versus at the top of the initial income distribution. And as in any observational study there is the possibility that our results are driven by some unobserved confounder, such as heterogeneity in local tastes for equality for example. We show, however, that our estimates are almost unchanged if we adjust them to account for potential selection on unobservables as proposed in Oster (Forthcoming). Together with our evidence on human and physical capital accumulation, our study thus provides reasonably well-identified estimates of the link between income inequality and growth.

The paper is organized as follows. Section 3.2 describes the Brazilian setting, and section 3.3 describes the data and presents summary statistics. Section 3.4 discusses our approach to analyze the relationship between left and right tail inequality and growth and how we deal with potential confounding factors. Section 3.5 presents and discusses our main results, and section 3.6 presents evidence on mechanisms. Section 3.7 presents the results of multiple robustness checks, and section 3.8 concludes with a discussion of external validity.

3.2 The Brazilian Setting

The starting point of our analysis is 1970, which is dictated by the availability of comparable income data over time as further discussed below. Our units of analysis are the 3,659 Brazilian *Áreas Mínimas Comparáveis* (AMCs), which are roughly equivalent to Brazil's municipalities in 1970. On average, AMCs had about 25 thousand inhabitants at that time. And while today Brazil is a middle-income country with a large urban population, this was by no means true in 1970 when a large fraction of its population lived in poverty, and more than half resided in rural areas. The level of education was also extremely low. In particular, AMCs had an average educational attainment of individuals above 25 years old of only 1.37 years, and an iliteracy rate of 44% for people above 15.

Since 97% of individuals who worked or studied did so in their municipality of residence, several mechanisms driving the relationship between income inequality and growth should operate within AMCs. While there was ample room for growth driven by the accumulation of human and physical capital, opportunities for investment were rather limited for households at the bottom of the wealth distribution since access to credit was not widespread. For example, in the agricultural sector where 42% of the workforce was employed in 1970, only 12% of establishments received credit during that year.[2] Credit constraints were therefore likely binding for a large fraction of Brazilians in our period of study. Together with setup costs, inequality at the lower or upper tail of the income distribution might therefore lead to very different growth dynamics as argued in Galor and Zeira (1993), for example.

An important part of the literature has devoted attention to the role of political forces in explaining the relationship between inequality and economic development (e.g., Persson and Tabellini, 1994; Alesina and Rodrik, 1994; Benabou, 1996, among others). In particular, these studies posit that more unequal societies face higher pressure for redistribution, which in turn generates distortions and hampers growth. Although Brazil was under a military dictatorship from 1964 until 1985, local elections were still held in most municipalities. Furthermore, while only 2.6% of total revenues were raised by municipal taxes, around 12%-17% of total public spending was done by municipal governments (Hagopian, 1996). So even though within-AMC inequality in the 1970s could not impact local taxation in a relevant way, it might still affect the composition of spending and thus economic development. While we do not mean to play down the role of redistributive policies in mediating the effect of local inequality on subsequent growth, it is not clear from a theoretical perspective how inequality generated at the bottom as opposed to the top of the income distribution would interact with spending decisions at the AMC level. Furthermore, lack of information on spending at the local level for this period does not allow us to explore this issue further.

[2]These figures are based on our own calculations using the 1970 population and agricultural censuses. Information on access to credit for households or firms in other sectors is not available for this period.

3.3 Data and Descriptive Statistics

Our analysis relies on the 25% sample of the 1970 and 1980 Brazilian censuses obtained from the Brazilian Statistical Agency (*Instituto Brasileiro de Geografia e Estadística*, IBGE), and on AMC-level statistics published by IPEA (*Instituto de Pesquisa Econômica Aplicada*).[3] The starting point of our analysis is 1970 since this is when the first round of the Brazilian census with precise information on individual incomes was conducted.[4] Our units of analysis are the 3,659 Brazilian AMCs, which are themselves based on all existing municipalities from 1970 to 2000. Since many municipalities split or merged with others after 1970, doing our analysis at the AMC-level allows us to keep the borders constant and follow the same geographical units over time.[5]

When working with the 1970 25% census sample we first match the 3,974 municipalities appearing in this census to their corresponding AMCs.[6] This census investigated the monthly income for all individuals 10 years and older and asked for: (i) the income of the last month for those who earn a fixed income (e.g., salaries, pensions, etc.); (ii) the average monthly income in the last twelve months for those who receive variable income (e.g. professionals' fees, sale and brokerage commissions, payments for services rendered, etc.); and (iii) the monthly average of other regular sources of income such as routine donations, rents, dividends, etc. We construct the per capita family income distribution for each AMC in 1970 by dividing the sum of the individual incomes of all family members living in the same household by the number of family members.[7] This way, all family members living under the same roof have the same per capita income. We exclude from our analysis those individuals living in collective dwellings (e.g. hotels, hospitals, nursing homes), which amount to 1.89% of our sample. We also exclude individuals living in a private dwelling who are unrelated to the family head (tenants and domestic servants) and who account for 2.21% of all individuals. We

[3]Available at http://www.ipeadata.gov.br/.

[4]In the previous census round in 1960, income was reported in only eight categories.

[5]Brazil had 3,974 municipalities in 1970, and 5,507 by 2000.

[6]We match municipality and AMC codes using the Data Zoom program developed by the Department of Economics at PUC-Rio, available at http://www.econ.puc-rio.br/datazoom/english/.

[7]Only 1.68% of individuals who report having a source of income do not report their earnings.

then construct three main indicators from each AMC's per capita family income distribution, using the appropriate expansion weights provided by IBGE. First, we calculate the average per capita family income in 1970, which we express in R$ of 2000. Second, we construct the 1970 AMC Gini coefficient,[8] and third the share of total AMC income held by each of the quintiles. We also calculate an approximation to the Gini coefficient using these quintiles shares.[9] Unlike subsequent censuses, incomes above Cr$ 9,998 are top-coded at this value,[10] affecting 0.04% of employed individuals. As a robustness check, we adjust top-coded incomes, multiplying them by a factor of 2.15 so that individual incomes in the top 20% follow a Pareto distribution.[11] We also use the 1970 census to compute the share of occupied individuals working in each of the 16 economic sectors detailed in the census, which we use as controls in the robustness checks we perform in Section 3.7.1.[12]

We apply the same procedure to the microdata from the 1980 25% long-form sample, in order to obtain the per capita family income distribution of each AMC, and then calculate different per capita income percentiles for each AMC. For computing poverty rates, we use three different poverty lines: (i) half of the Brazilian minimum wage in September 1991; (ii) US$ 2 a day at 2005 PPP, which is the median poverty line amongst developing countries based on a compilation of national poverty lines in Ravallion et al. (2009); and (iii) US$ 1.25 a day at 2005 PPP, the mean poverty line for the poorest 15 countries. The first of these was obtained from IPEA, whereas the others were taken from Ravallion (2012). We also rely on the 1980 census 25% sample to study the migration patterns across AMCs between the 1970 and 1980 censuses. More specifically, we compute immigration

[8] We use the *ineqdec0* code written by Stephen Jenkins for this calculation.

[9] Define Qn as the share of total AMC income held by quintile n. Then Gini $\approx 0.8 \times$ [Q5+0.5Q4−0.5Q2−Q1].

[10] All figures in the 1970 and 1980 census are reported in Cruzeiros (Cr$), Brazil's currency at the time. We converted all figures to Brazilian Reais (R$) of 2000 using the guidelines employed by the 1998 "Atlas de Desenvolvimento Humano no Brasil."

[11] This methodology is commonly used by researchers working with CPS data in the US. Examples include Katz and Murphy (1992), Autor et al. (2008), and Autor and Dorn (2013).

[12] These sectors are agriculture and forestry; gathering of wild growing products; hunting and fishing; mining and quarrying; manufacturing; construction; public utilities; wholesale and retail trade; services; transporting and communications; education, health and social activities; public administration, legislation and justice; national defense and public safety; real estate, financial and insurance activities; liberal professions; and other activities.

and emigration rates for each AMC between 1970 and 1980. Since the 1980 census asks individuals how long they have been living in their current municipality, we count all individuals in a particular AMC who report that they were not living in their current municipality in 1970 as immigrants. Individuals who are younger than 10 years old in 1980 and belong to a family in which the head is an immigrant are considered immigrants as well. We calculate an AMC's immigration rate as the ratio between the number of immigrants in 1980 and the AMCs total population in 1970. Furthermore, the census also asks people who have been living in their current municipality for less than 10 years to specify the municipality in which they were previously residing. Thus, for each AMC, we can calculate the number of people who were living there in 1970 and left. We use this to calculate the emigration rate, which is simply the number of emigrants of an AMC divided by the 1970 population. A caveat for this measure is that the municipality of origin is missing for approximately 19% of all immigrants. Since we cannot trace these people to their municipality of origin, our emigration rate does not include these observations in the numerator. We also calculate AMCs' fertility and mortality rates, to uncover the population dynamics in this period. The fertility rate is the ratio between the number of AMC natives who are less than 10 years old in 1980 and the population in 1970. The mortality rate is therefore the ratio between the change in population between 1970 and 1980 not accounted for by fertility and migration, divided by 1970 population.[13]

From IPEA we obtain the following 1970 AMC-level control variables, which we use in all our regressions: average years of schooling of individuals aged 25 and above, literacy rates for people 15 years and older, total population, the percentage of people living in urban areas and life expectancy. We also obtain a set of time invariant AMC-level controls such as latitude, longitude, distance to the state and federal capitals, and an indicator for whether the AMC is located on the coast. For outcomes we got the mean per capita family income at the AMC-level for 1980, 1991 and 2000, which was itself based on the corresponding population censuses. We calculate our main outcome variable, the growth in AMC mean family income per capita, as the log difference in real mean per capita family in-

[13] This also includes individuals who emigrated from an AMC but do not report their municipality of origin.

come between 2000 and 1970. We also obtained several measures of educational attainment in 1980 at the AMC-level, which we also use as outcomes, and which are based on the educational level of individuals 25 years and older. Specifically we use average years of education, and the share of individuals with less than 4, between 4 and 8, and with 8 or more years of education for each AMC.

Other IPEA data include the value of capital stocks held by businesses in each AMC in 1970 and 1980 in the agricultural, commercial, manufacturing and service sectors, all based on the respective economic censuses.[14] Up until 1980, Brazil's statistical agency carried out periodic economic censuses covering all firms in each of these sectors. As explained in detail by the academics in charge of performing these calculations at IPEA (Reis et al., 2005), when calculating the value of capital stocks for agricultural establishments they include farmland, buildings, long-term crops,[15] vehicles, machinery, agricultural instruments, and livestock. They deduct the value of residential buildings within farms, and only consider livestock used for traction or reproduction. Firms in the agricultural sector include all establishments dedicated to farming, cattle, poultry or rabbits, beekeping, raising silk worms, horticulture, floriculture, forestry, and extraction of vegetable products. When calculating the value of capital stocks for manufacturing, commercial and service industry establishments, they take into account the value of firms' capital employed in buildings, land, machinery and equipment as reported in the corresponding economic censuses. The firms covered by the commercial census are all the establishments dedicated to the purchase, sale, exchange or distribution of merchandise through retail.[16] Activities considered in the manufacturing census include the processing and packaging of food products, metallurgical activities, production of pharmaceutical products, clothes items, etc. Finally, firms in the service sector include all establishment whose activity in-

[14] A detailed account on how the value of capital stocks at the AMC-level was backed out from the corresponding economic censuses by IPEA can be found in Reis et al. (2005).

[15] Long-term crops are those that do not need to be replanted after each harvest, such as coffee, oranges, bananas, etc.

[16] For example, the sales activities of a firm that produces machinery is accounted for in the commercial census only if the firm sells its products through its own retail establishments, but not if it does so through a wholesaler. Further explanations can be found in the reports by IBGE on the commercial censuses. For example, at http://biblioteca.ibge.gov.br/visualizacao/periodicos/63/cc_1980_v4_n15_ba.pdf

volves providing services to people, such as hotels, repair shops, restaurants, and so on.[17] After calculating the value of each establishment's capital stocks, IPEA aggregates these figures at the municipality level, separately for each sector. In performing this calculation, they consider an establishment as belonging to a municipality if it is located there. As with all of our income figures, capital stocks are expressed in real terms (in 2000 R$). We compute the growth in the value of AMC capital stocks between 1970 and 1980 as the log difference in the real value of capital stocks in all sectors of the economy.

We summarize the main variables for our analysis in Table 3.1. In 1970 Brazil was an extremely poor country. The average AMC monthly mean per capita family income in 1970 was 56 R$ (in R$ of 2000), which is approximately 38 US dollars as of March 2016. Inequality rates were high with an average Gini coefficient of 0.47 and standard deviation 0.07. Inequality also displayed a considerable degree of spatial variation across AMCs, as shown in Figure 3.2. During the 1970-2000 period income per capita more than doubled on average across AMCs. Most of these gains occurred in the first decade and were accompanied by large increases in physical capital stocks across sectors.

3.4 Estimation Approach

In order to estimate the effect of initial overall income inequality on subsequent economic growth, we run the following OLS regression:

$$\begin{aligned} ln(\overline{y}_{a,s,t}) - ln(\overline{y}_{a,s,1970}) = \beta_0 + \beta_1\, ln(\overline{y}_{a,s,1970}) + \beta_2\, Gini_{a,s,1970} \\ + X_{a,s,1970}\, \delta + \gamma_s + U_{a,s,t} \end{aligned} \tag{3.1}$$

where $\overline{y}_{a,s,t}$ is the mean per capita family income in AMC a in state s and year t. We estimate this regression with growth in AMC mean per capita family income in 1970-2000 as our outcome variable (i.e., when t is 2000). In order to assess the

[17]Further details can be found in IBGE's reports on the results of the service industry census. See http://biblioteca.ibge.gov.br/visualizacao/monografias/GEBIS%20-%20RJ/censodosservicos/1980_v05_n03_AC.pdf

effect of initial inequality on future levels of income per capita directly, we also use the same specification but replace income growth with the (natural) logarithm of average per capita family income in 1980, 1991 and 2000, $ln(\overline{y}_{a,s,t})$, as the outcome variable. $Gini_{a,s,1970}$ is the Gini coefficient in AMC a in state s in 1970, $X_{a,s,1970}$ is a vector of 1970 AMC-level controls, γ_s are state fixed effects and $U_{a,s,t}$ is the influence of unobserved factors on outcomes in year t.[18]

Our coefficient of interest in these regressions is β_2, the effect of initial inequality on the long-run level or growth rate of income per capita. There are many potential confounders at the AMC-level in 1970 that could correlate with both initial income inequality and subsequent economic growth, and the direction of the bias in $\hat{\beta}_2$ is unclear. For instance, AMCs with greater income inequality in 1970 might also be places where a higher percentage of the population has low levels of education, and low education is likely bad for economic growth, biasing $\hat{\beta}_2$ downwards. AMCs with high inequality in 1970 might also be more rural, and growth patterns of rural areas might be different from those of more urbanized AMCs for reasons unrelated to the society's initial income inequality. We address potential omitted variable bias by including standard growth determinants in all our regressions as well as state fixed effects.[19] In particular, $X_{a,s,1970}$ includes a set of AMC characteristics in 1970 (average years of schooling, literacy rate, population, % of urban population, and life expectancy), as well as other time invariant features of each AMC (latitude, longitude, distance from the federal and state capitals, and an indicator for whether the AMC is located on the coast). The key assumption for causal interpretation of $\hat{\beta}_2$ is that unobserved determinants of future income per capita are mean-independent of inequality, conditional on initial income per capita and our controls. While this assumption is not directly testable, we think it is a reasonable assumption, given our supporting evidence on mechanisms and

[18]Controlling for AMC fixed effects would require us to find valid instruments to deal with the presence of the lagged outcome. We prefer the simplicity and transparency of standard regression control. Moreover, as pointed out by Easterly (2007) for example, it is not clear whether using relatively high-frequency data is appropriate for the question we wish to study since income inequality tends to be fairly stable over time and its impact on subsequent growth is believed to unfold over decades, if not generations. Our estimates are consistent with this notion: although the effect of initial income inequality on growth is strongest in the first decade, it takes around 20 years for the full effect to materialize

[19]Excluding *Distrito Federal* which is also a municipality in itself, Brazil has 26 states in total.

the fact that our controls absorb a large part of the variation in outcomes as shown below.

In order to distinguish between effects of inequality originating from either tail of the income distribution, we take advantage of the fact that the Gini coefficient can be approximated with a formula based on the shares of income held by each of the quintiles. That is:

$$Gini_{a,s,1970} \approx 0.8 \times [Q5_{a,s,1970} + 0.5Q4_{a,s,1970} - 0.5Q2_{a,s,1970} - Q1_{a,s,1970}] \tag{3.2}$$

where $Qn_{a,s,1970}$ is the 1970 share of total income of AMC a in state s held by quintile n. As can be seen in the first column of Table 3.2, controlling for state fixed effects and our vector of 1970 AMC covariates, the Gini coefficient and its approximation based on quintile shares in 1970 vary almost one-to-one, with an R-squared of almost one. In light of this, decomposing differences in the 1970 AMC Gini coefficients into differences in quintile income shares as in (3.2) allows us to differentially focus on the growth effects of inequality in the left and right tails of the income distribution. Throughout our Gini decomposition exercise, the omitted quintile is the middle one. Thus, a decrease in the income held by the first quintile implies an increase in the income held by the middle one, and a higher overall income inequality, as illustrated in Panel A of Figure 3.1. Throughout the paper we refer to this as inequality in the left or bottom tail. The exact same increase in overall inequality occurs when a higher percentage of overall income is held by the top quintile at the expense of the middle one. This is what we call higher inequality in the right or upper tail, as shown in Panel B of Figure 3.1. With this intuition in mind, we distinguish between the growth effects of inequality in the left and right tails by running the following regressions:

$$\begin{aligned} ln(\overline{y}_{a,s,t}) - ln(\overline{y}_{a,s,1970}) = \beta_0 &+ \beta_1\, ln(\overline{y}_{a,s,1970}) + \alpha_1\, Q5_{a,s,1970} + \alpha_2\, Q4_{a,s,1970} \\ &+ \alpha_3\, Q2_{a,s,1970} + \alpha_4\, Q1_{a,s,1970} + X_{a,s,1970}\,\delta \\ &+ \gamma_s + U_{a,s,t} \end{aligned} \tag{3.3}$$

which is the specification in (3.1), but replacing the Gini coefficient with four of the quintile income shares and omitting the middle quintile share. In this regression, our coefficients of interest are α_1 (the coefficient for inequality in the right tail), and α_4 (the coefficient for inequality in the left tail, when multiplied by minus 1). When exploring the correlation between inequality in the left and right tails with subsequent growth in physical capital, we run (3.3) with the real value of the aggregate capital stock held by firms instead of per capita income. We do this separately for each sector of the economy (agriculture, manufacturing, commerce and services), and also for the total capital stock across sectors. In all these regressions we control for the log of the 1970 value of the capital stocks held by firms in all sectors (as well as the log of the respective sector-specific capital stocks in 1970).

When analyzing growth in human capital, we run the above regression for a set of outcomes capturing the 1980 levels educational attainment in an AMC, such as average years of education of individuals above 25 years old, the percent of such individuals with less than 4 years of education (i.e, less than a primary school degree), between 4 and 8 years (i.e., more than primary but less than middle school), and 8 or more years of education (i.e., at least a middle school diploma). In addition to the baseline controls included in $X_{a,s,1970}$, we also control for the 1970 proportion of individuals 25 and older with less than 4, between 4 and 8, and 8 or more years of education.

3.5 Results

3.5.1 Overall Inequality and Income per Capita Growth

Column 1 of Table 3.3 shows that there is positive correlation between the Gini coefficient in 1970 and growth in the period spanning 1970-2000. AMCs with a Gini that was one standard deviation (0.07 Gini points) higher in 1970 grew about 3% more between 1970-2000. The results are very similar when using the Gini approximation based on quintile shares, as can be seen in column 2, lending credibility to the regressions based on equation (3.3) below. Moving to the regressions with the log of income per capita in 1980, 1990 and 2000 as outcomes, it is clear from columns 3 to 5 that the results are much stronger in the first decade. Income per capita increased by about 2% by 1980 and by about 3% by 1991 in places where the Gini coefficient was 7 percentage points higher, with only negligible additional growth by the year 2000. Taking these results together, we conclude that AMCs with higher inequality in 1970 end up with higher average income in 2000, but do not experience permanently higher growth.

In line with the results from cross-country growth regressions, the coefficient for the income lag is negative and statistically significant in the first two columns, meaning that AMCs that start out with a higher income level grow at a slower rate. Even though our study explores within-country (across sub-national unit) variation, Brazilian AMCs also experience income convergence as predicted by growth theories, which speaks to the external validity of our study and suggests that there are at least some common mechanisms linking inequality and growth both within and across countries. We also note that our regressions account for most of the variation in subsequent income per capita levels, (R-squared of 0.877 in column 5 of Table 3.3), leaving little room for unobserved confounders to dramatically alter our estimate of interest.

3.5.2 Quintile Income Shares and Income per Capita Growth

Having established a positive correlation between inequality as measured by the Gini coefficient in 1970 and subsequent economic growth, we now explore whether this effect is different when inequality originates from the lower as opposed to the upper tail of the income distribution. As explained in Section 3.4, the omitted quintile in regressions with quintile shares is the third quintile. Thus, a *decrease* in the income held by the first quintile is matched by an equivalent increase in the income share of the middle quintile, implying *higher* inequality in the left tail. Therefore, multiplying the coefficient associated with the share of 1970 AMC income held by Q1 by -1 gives us the the effect of an increase in left-tail inequality on growth of AMC income per capita. On the other hand, an *increase* in the percentage of income held by the top quintile implies a decrease in the percentage of income held by the middle quintile, and an *increase* in inequality in the right tail. Therefore, the coefficient on the share of 1970 AMC income held by Q5 directly gives the partial effect of an increase in inequality in the right tail on growth of AMC income per capita.

As shown in the first column of Table 3.4, AMCs with higher inequality in the right tail (i.e., a higher share of income held by Q5 at the expense of Q3) did not grow more in 1970-2000. On the other hand, the negative coefficient for the share of income held by Q1 means that AMCs with higher inequality in the left tail of the distribution did experience higher growth. In particular, AMCs with a 1970 Q3 income share higher by one standard deviation (3 percentage points) at the expense of Q1 grew about 3% more over the period 1970-2000. Income per capita increased by about 4% by 1980 with little additional growth by 1991 and a slight and statistically insignificant drop by 2000. As with overall inequality, higher left-tail inequality does not lead to a permanent increase in income per capita growth. The last row of Table 3.4 shows that the differential impact between left- and right-tail inequality is not only economically but also statistically significant in most specifications. We therefore conclude that the overall effect of inequality picked up by the Gini coefficient is entirely driven by the lower tail of the initial income distribution: compared to more equal places, AMCs with a higher share of income going to the middle quintile at the expense of the bottom quintile grow more rapidly, while places with a higher share of income going to the top quintile at the expense of the middle quintile get no growth boost at all.

3.5.3 Quintile Income Shares and Subsequent Income Distribution

In this subsection, we analyze whether the higher growth in per capita income experienced by AMCs with greater initial inequality in the left tail had any distributional impacts. Since this higher growth already materializes by 1980, we focus on this period. As displayed in Table 3.5 and Figure 3.3, more inequality in the left tail in 1970 is correlated with a positive shift in the top half of the AMC's income distribution. More specifically, in AMCs in which the share of income held by Q3 (Q1) was 3 percentage points higher (lower) in 1970, the per capita income percentiles in the top half of the distribution were between 4% and 6% higher in 1980.

Consistent with these results, Table 3.6 shows that higher initial inequality in the left tail is associated with significantly lower poverty rates, but only for broad definitions of poverty. Under our two broadest definitions, for which the average poverty rates were 60% and 45%, AMCs in which the share of income held by Q3 in 1970 was 3 percentage points higher (and the share of Q1 was lower) had a poverty rate of about 1 percentage point lower in 1980. Higher initial inequality does not correlate with lower poverty rates in 1980 for our strict definition of poverty.

3.6 Evidence on Mechanisms

3.6.1 Quintile Income Shares and Physical Capital Growth

Given the positive correlation between 1970 inequality in the left tail and subsequent growth in mean per capita family income, we should observe a similar correlation with growth in physical and human capital if credit constraints and setup costs are important. In Table 3.7, we regress real growth in the value of the capital stocks held by firms from different sectors in the period spanning 1970-1980 against the quintile income shares, the 1970 value of both total and sector-specific capital stocks in 1970 and our other controls. Consistent with the zero growth effect of right-tail inequality discussed above, we find small and statistically insignificant effects of inequality in the right tail on growth in firms' capital stocks for three out of four sectors as well as overall. On the other hand, we find a positive and sizable correlation between inequality in the left tail in 1970 and growth in the value of capital stocks from 1970 to 1980 for all four sectors as well as overall. Total capital stocks grew about 10% more in real terms in AMCs in which the 1970 income share of Q3 was one standard deviation higher at the expense of the bottom quintile. The effect of left-tail inequality on physical capital accumulation arises across sectors, ranging from about 9% in agriculture, to about 16% in the commercial sector, about 27% in manufacturing and about 15% in services.

3.6.2 Quintile Income Shares and Human Capital Growth

Turning to investments in education, Table 3.8 shows a similar pattern. The first column documents that there is no correlation between initial inequality in the right tail and growth in average educational attainment over the 1970-1980 period. On the other hand, average years of schooling experienced higher growth in 1970-1980 in AMCs that started out with higher inequality in the left tail. Though significant statistically, this effect is relatively small: AMCs in which the income held by the middle quintile in 1970 was 3 percentage points higher (at the expense of the bottom quintile) saw an increase of 0.03 years in average educational attainment of individuals above 25 years of age. Turning to the regressions in columns 2 to 4, it is clear that this increase in educational attainment was driven by a smaller proportion of the population with less than 4 years of education (i.e., less than a primary school degree), and a higher proportion with educational attainment of between 4 and 8 years. Higher inequality at the top also increased the proportion of the population with more than 8 years of schooling, but the impact is negligibly small.

3.6.3 Quintile Income Shares and Migration

A plausible concern is that our results are driven by differential migration patterns. For example, AMCs with high initial left-tail inequality could be attracting workers with higher education and higher potential earnings, leading to a selection-driven increase in average income. As shown in column 1 of Table 3.9, AMCs with a lower Q1 income share and higher share of income held by the third quintile exhibit higher population growth between 1970 and 1980, but the impact is very small. In particular, the population in AMCs with a 1970 Q3 income share higher by one standard deviation (3 percentage points) at the expense of Q1 grew about 4.5% more over the period 1970-1980. And as shown in column 2 of Table 3.9, this increase in population was not driven by immigration, since AMCs with higher inequality in the left tail in 1970 did not experience higher immigration between 1970 and 1980, the decade of highest growth. Places with high initial inequality in the left tail might experience higher out-migration of the poor and thus higher income per capita among remaining residents in future periods. As shown in column 3 of Table 3.9, we actually find the opposite: AMCs with higher inequality in the left tail in 1970 experienced lower emigration rates.[20] However, this effect is negligible in practice. A 3 percentage point increase in the share of total income held by the middle quintile (at the expense of the bottom quintile) is associated with an emigration rate 0.70 percentage points lower over the 1970-1980 decade relative to an average emigration rate of about 19%. It turns out that the higher growth in population between 1970 and 1980 experienced by AMCs that started out with higher inequality in the left tail is mostly driven by a lower mortality rate, as shown in column 5 of Table 3.9.[21]

3.7 Robustness Checks

3.7.1 Controlling for 1970 Sectoral Labor Force Shares

While our main specification controls for the share of an AMC's 1970 population living in rural areas, as a robustness check we also account for differences in the initial structure of the economy in a more flexible manner. In particular, we control for the share of occupied individuals working in each of the 16 economic sectors defined by the 1970 census, as detailed in Section 3.3. As shown in Tables 3.A1 and 3.A2, the association between inequality in 1970 and subsequent economic growth is robust to the inclusion of these controls. Our evidence on channels featured in Tables 3.A3 and 3.A4 is also consistent with our results on inequality, although slightly weaker when it comes to physical capital accumulation.

3.7.2 Imputing Top-Coded Incomes

Unlike subsequent censuses, incomes in the 1970 census are top-coded, a practice which affects 0.04% of employed individuals. In order to check whether our results are driven by differential measurement error at the bottom versus at the top

[20] These results should be taken with care, since the municipality of origin is missing for approximately 19% of all immigrants.

[21] What we refer to as mortality rate is actually a residual category, namely the ratio between the change in population between 1970 and 1980 not accounted for by fertility and migration and the 1970 population. This includes not only people who passed away in 1970-1980, but also individuals who emigrated from the AMC but did not report their municipality of origin.

of the initial income distribution we impute top-coded incomes and construct new quintile shares. Following the methodology used by Katz and Murphy (1992), Autor et al. (2008), and Autor and Dorn (2013), among others, we multiply top-coded incomes by a factor of 2.15, so that individual incomes in the top 20% follow a Pareto distribution. As can be seen in Table 3.A5, our main results are robust to these imputations.

3.7.3 Alternative Definition of the 1970 Census Universe

As explained in Section 3.3, our main specification excludes individuals living in collective dwellings and individuals who live in a private dwelling but are unrelated to the family head (i.e., tenants and domestic servants), which in total account for 4.10% of individuals in the 1970 census. While the correlation between initial inequality in the left tail and subsequent growth in income per capita is robust to the inclusion of these individuals, as shown in Table 3.A6, inequality at the top is positively correlated with growth in income per capita in some specifications. However, the coefficients for inequality in the top are smaller and not robust across specifications.

3.7.4 Adjusting for Selection on Unobservables

As discussed in Section 3.4, there could be many confounders at the AMC-level in 1970 correlating with both initial income inequality and subsequent economic growth. Although we address potential omitted variable bias by including standard growth determinants in all our regressions as well as state fixed effects, we cannot fully rule out the existence of unobservable determinants of AMC growth which correlate with initial income inequality even conditional on these controls.

In this subsection, we follow the approach of Oster (Forthcoming), itself an extension of the methodology developed by Altonji et al. (2005), to evaluate the robustness of our estimates to potential omitted variable bias. Under the two assumptions that observable and unobservable variables are equally related to the regressor of interest and that the bias from unobservables is not so large that it biases the direction of the covariance between the observables and the regressor of interest, Oster (Forthcoming) develops an estimator that accounts for selection on unobservables. We also assume, as proposed in that paper, that the hypothetical maximum R^2 from a regression of the dependent variable against all observable and unobservable controls is the minimum value between 1 and 1.3 times the R^2 of the regression with observable controls. Since the quintile income shares only capture inequality in the left and right tails if they are conditioned on the other quintile shares and initial income, we include all of these in the "uncontrolled" regression.

Our estimates are almost unchanged if we adjust them to account for potential selection on unobservables. The bias-adjusted estimate for the Gini coefficient in the regression using growth in 1970-2000 as the outcome variable is equal to 0.402, down from 0.447 in the specification that controls for all our observables (column 1 in Table 3.3). Intuitively, the change is so small because the increase in R-squared is massive (from 0.013 in the uncontrolled regression to 0.673 controlling for observables) compared to the change in coefficient estimates (from an uncontrolled 0.541 to a regression-controlled 0.447). If we apply the same procedure to the first quintile income share (again with growth in 1970-2000 as the outcome variable), the adjusted estimate is -0.902, which is practically unchanged from the controlled estimate of -1.065 in column 1 of Table 3.4. In the case of Q5, which was small and statistically insignificant in our initial regression, controlling for potential omitted variable bias results in an impact estimate of 0.169, compared to 0.243 in column 1 of Table 3.4.

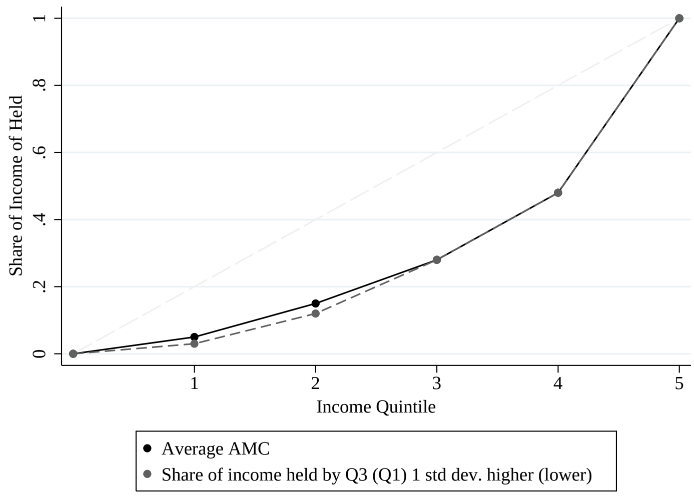
3.8 Conclusion

This study investigates whether inequality originating in the lower as opposed to the upper tail of the income distribution has different effects on subsequent income per capita growth. Using sub-national variation in Brazil, we find that holding average income per capita and an extensive set of controls constant, AMCs with higher inequality in the left tail of the income distribution in 1970 exhibited higher growth in income per capita over the subsequent three decades. At the same time there is no correlation between initial inequality in the right tail of the AMC income distribution and growth. We show that our estimates are remarkably robust when we account for selection on unobservables. Moreover, our results are barely affected if we flexibly control for 1970 structural differences across sectors, impute incomes that were top-coded in the 1970 census, or use alternative definitions of the population underlying our inequality measures. Consistent with the existence of credit constraints and setup costs for investing in physical and human capital, we show that AMCs that started out with higher inequality in the left tail also accumulated physical and human capital at a faster pace while right-tail inequality has no such effects.

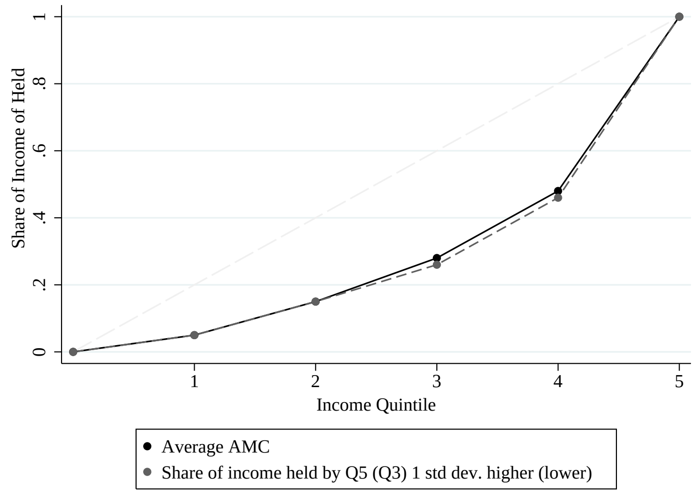
Whether higher inequality at the bottom of the income distribution would lead to higher growth in other contexts is not obvious. One could easily think of situations in which increasing inequality at the bottom of the income distribution would have a neutral impact on growth, or could even hamper it. Consider once more a stylized economy in which the population is divided into three groups of equal size (the poor, the middle class, and the rich). If the incomes of the poor and the middle class are initially too low to overcome the setup costs for investing in either human or physical capital, an increase in inequality at the bottom while keeping overall income constant (i.e., a transfer of income from the poor to the middle class) might allow the middle class to overcome the setup costs and invest in human and physical capital. But consider instead an economy where credit constraints only bind for the poorest group. Higher inequality in the left tail would have no impact on growth in this situation. In an even richer economy in which all groups can profitably invest in human and physical capital, higher inequality in the lower-tail could even be bad for growth if it results in the poor becoming credit constrained. Nonetheless, distinguishing whether inequality originates from the lower or upper tail of the distribution (or both) may provide an indication as to whether credit constraints and setup costs are part of the explanation for whatever overall relationship between inequality and growth is found in another setting.

Figure 3.1: Gini Coefficient and Inequality at the Top and Bottom of the Income Distribution

Panel A: increase in overall inequality originating from the left-tail of the distribution

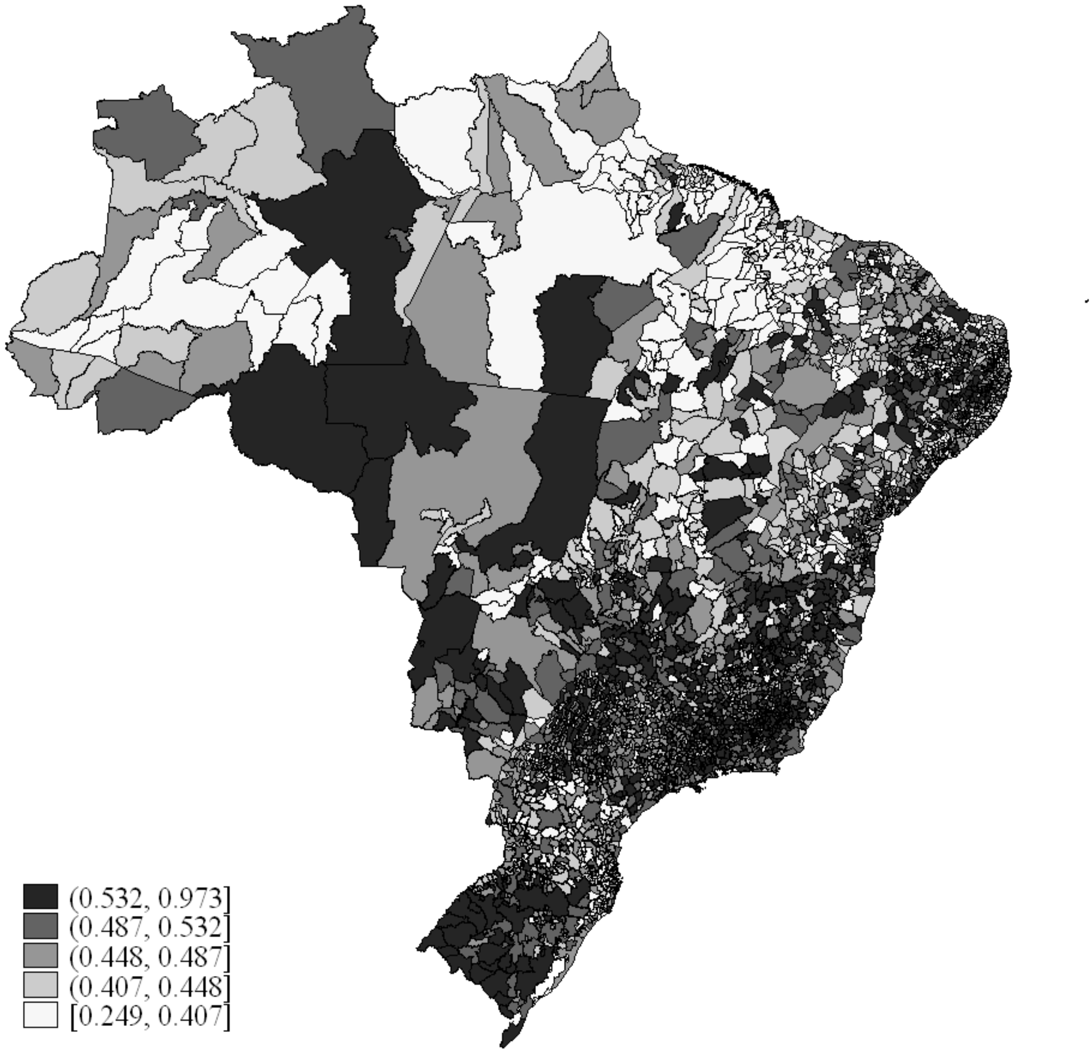


Panel B: increase in overall inequality originating from the right-tail of the distribution



Notes: The solid line displays the Lorenz curve of an AMC with the average income held by each of the quintiles. In the top figure, the dashed line shows the Lorenz curve of the average AMC when the % of income held by the third quintile is increased by 1 standard deviation (3 percentage points), at the expense of the first quintile (i.e., an increase in overall inequality generated from the left-tail of the income distribution). In the bottom figure, the dashed line shows the Lorenz curve of the average AMC when the % of income held by the fifth quintile is increased by 1 standard deviation (3 percentage points), at the expense of the middle quintile (i.e., an increase in overall inequality generated from the right-tail of the income distribution.)

Figure 3.2: Gini Coefficient Across Brazilian AMCs in 1970



Notes: Each unit is an Área Mínima Comparável (AMC) in 1970. Darker areas indicate greater income inequality as measured by the Gini coefficient in 1970.

Table 3.1: Descriptive Statistics

	Mean	Std. Dev	Min	Max
Dependent Variables				
Ln (real mean per capita family income)				
1970-2000 growth (log difference)	1.13	0.35	-0.75	3.67
1980 mean (in 2000 R$)	4.77	0.59	2.32	6.41
1991 mean (in 2000 R$)	4.64	0.59	3.17	6.38
2000 mean (in 2000 R$)	5.01	0.57	3.62	6.86
Growth in aggregate capital stocks in 1970-1980				
Agriculture	1.32	0.70	-2.99	18.72
Commerce	1.28	1.17	-13.02	10.41
Manufacturing	1.66	2.47	-12.55	20.36
Services	2.00	2.11	-12.62	16.59
Total	1.39	0.61	-1.40	7.09
1980 educational attainment (people 25 years and older)				
Average years of schooling	2.07	1.06	0.10	7.20
Proportion with less than 4 years of schooling	0.74	0.16	0.15	0.99
Proportion with 4 or more and less than 8 years of schooling	0.20	0.12	0.01	0.75
Proportion with 8 or more years of schooling	0.06	0.05	0.00	0.48
Explanatory Variables - all measured in 1970				
Gini coefficient	0.47	0.07	0.25	0.97
Gini approximation based on quintile income shares	0.42	0.07	0.15	0.80
Share of AMC income held by Q1	0.05	0.02	0.00	0.14
Share of AMC income held by Q2	0.09	0.02	0.00	0.30
Share of AMC income held by Q3	0.14	0.03	0.00	0.27
Share of AMC income held by Q4	0.20	0.03	0.00	0.45
Share of AMC income held by Q5	0.52	0.07	0.14	1.00
Ln (real mean per capita family income) (2000 R$)	3.89	0.54	0.57	5.70
Average years of schooling (25 years and older)	1.37	0.81	0.00	5.60
Iliteracy rate (15 years and older)	0.44	0.18	0.03	0.92
Population (in 000s)	25.45	132.47	0.83	5924.61
Proportion of residents living in urban areas	0.33	0.21	0.01	1.00
Life expectancy	51.11	4.27	38.40	64.46

Notes: The unit of observation is an Área Mínima Comparável (AMC) over the period 1970-2000. There are 3,659 AMCs.

Table 3.2: Income Shares and Income Inequality in 1970

Dependent variable: 1970 Gini coefficient		
Gini approximation	1.099***	
	(0.006)	
Share of 1970 AMC income held by Q5		0.925***
		(0.015)
Share of 1970 AMC income held by Q4		0.350***
		(0.023)
Share of 1970 AMC income held by Q2		-0.332***
		(0.026)
Share of 1970 AMC income held by Q1		-0.696***
		(0.029)
Observations	3,659	3,659
R^2	0.976	0.979

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC, and the dependent variable in both columns is the 1970 Gini coefficient. The explanatory variable of interest in column 1 is the 1970 Gini approximation based on quintile shares, calculated using the formula in (3.2). The explanatory variables of interest in column 2 are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 mean per capita family income, average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast.

Table 3.3: Income Inequality in 1970 and Subsequent Economic Growth

	1970-2000 growth		Ln (Income)		
	Ln difference		1980	1991	2000
Gini coefficient	0.447***		0.313***	0.415***	0.447***
	(0.060)		(0.073)	(0.072)	(0.060)
Gini approximation		0.521***			
		(0.066)			
Ln (1970 income)	-0.771***	-0.770***	0.386***	0.296***	0.229***
	(0.017)	(0.017)	(0.023)	(0.021)	(0.017)
Observations	3,659	3,659	3,659	3,659	3,659
R^2	0.673	0.673	0.857	0.849	0.877

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variable in columns 1 and 2 is the 1970-2000 growth in the mean per capita family income, and the dependent variable in columns 3, 4 and 5 is the mean per capita family income in 1980, 1990 and 2000 (in ln). The 1970 Gini approximation based on quintile shares is calculated using the formula in (3.2), and Ln (1970 income) is the mean per capita family income in 1970 (in ln). All regressions include state fixed effects and control for average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast.

Table 3.4: Income Shares in 1970 and Subsequent Economic Growth

	1970-2000 growth	Ln (Income)		
	Ln difference	1980	1991	2000
Share of 1970 AMC income held by Q5	0.243	0.183	0.092	0.243
	(0.170)	(0.226)	(0.198)	(0.170)
Share of 1970 AMC income held by Q4	0.216	0.126	-0.026	0.216
	(0.229)	(0.328)	(0.255)	(0.229)
Share of 1970 AMC income held by Q2	-0.349	0.359	-0.125	-0.349
	(0.277)	(0.336)	(0.320)	(0.277)
Share of 1970 AMC income held by Q1	-1.065***	-1.440***	-1.591***	-1.065***
	(0.380)	(0.469)	(0.453)	(0.380)
Observations	3,659	3,659	3,659	3,659
R^2	0.674	0.858	0.849	0.877
P-value (Q4+Q2=0)	0.746	0.362	0.748	0.746
P-value (Q5+Q1=0)	0.110	0.055	0.014	0.110

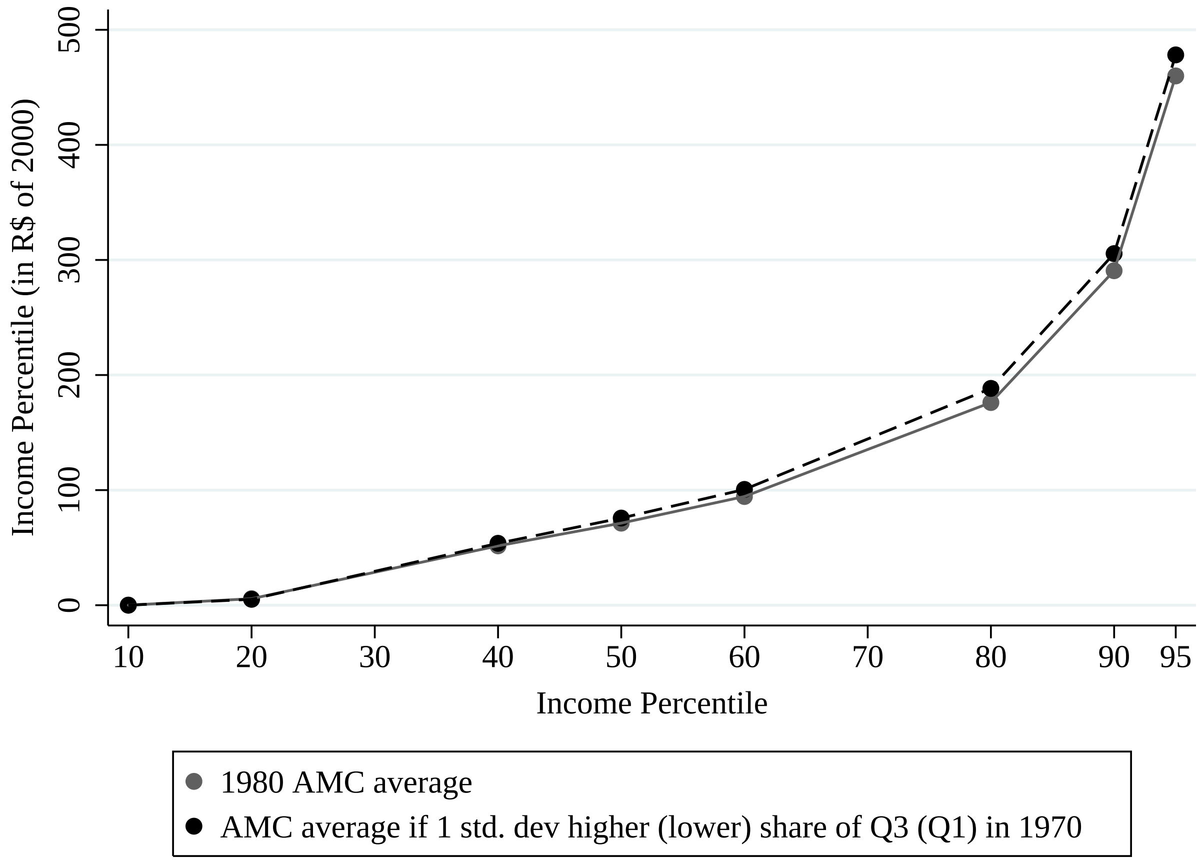
Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variable in column 1 is the 1970-2000 growth in the mean per capita family income, and the dependent variable in columns 3, 4 and 5 is the mean per capita family income in 1980, 1990 and 2000 (in ln). The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 mean per capita family income, average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast.

Table 3.5: Income Shares in 1970 and 1980 Income Percentiles

	Ln (Income Percentile in 1980)							
	10^{th}	20^{th}	40^{th}	50^{th}	60^{th}	80^{th}	90^{th}	95^{th}
Share of 1970 AMC income held by Q5	-0.038	0.262	-0.421	-0.576	-0.790**	-0.655*	-0.123	0.382*
	(0.028)	(0.688)	(0.455)	(0.423)	(0.382)	(0.341)	(0.196)	(0.216)
Share of 1970 AMC income held by Q4	-0.043	0.713	0.278	0.097	-0.148	-0.407	0.050	0.339
	(0.065)	(0.921)	(0.662)	(0.626)	(0.570)	(0.593)	(0.274)	(0.305)
Share of 1970 AMC income held by Q2	-0.067	3.425***	1.160	0.937	0.342	0.351	0.101	0.134
	(0.051)	(1.165)	(0.827)	(0.727)	(0.691)	(0.312)	(0.295)	(0.329)
Share of 1970 AMC income held by Q1	-0.084	2.986**	-1.429	-2.056**	-2.173***	-2.316***	-1.707***	-1.325***
	(0.060)	(1.482)	(0.965)	(0.829)	(0.663)	(0.598)	(0.425)	(0.474)
Observations	3,658	3,658	3,658	3,658	3,658	3,658	3,658	3,658
R^2	0.807	0.415	0.665	0.726	0.791	0.856	0.867	0.854

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variables are the different AMC income percentiles in 1980 (in ln), based on per capita family incomes. The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 mean per capita family income, average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast.

Figure 3.3: Inequality in the Left in 1970 and 1980 Income Percentiles



Notes: The solid line plots the 1980 income percentiles of the average AMC in terms of per capita family income. The dashed line plots the 1980 income percentiles of an AMC with a 1 standard deviation higher (lower) share of income held by the third (first) quintile in 1970, which was calculated using the coefficients in Table 3.6.

Table 3.6: Income Shares in 1970 and 1980 Poverty Rates

	% of people under poverty line in 1980		
	1/2 the Sep-91 min. wage (84.73 R$ a month)	US$ 2 a day (50.67 R$ a month)	US$ 1.25 a day (26.43 R$ a month)
Share of 1970 AMC income held by Q5	0.158***	0.084	0.014
	(0.057)	(0.061)	(0.062)
Share of 1970 AMC income held by Q4	0.004	-0.037	-0.070
	(0.073)	(0.082)	(0.088)
Share of 1970 AMC income held by Q2	-0.188**	-0.205**	-0.140
	(0.090)	(0.104)	(0.102)
Share of 1970 AMC income held by Q1	0.417***	0.326**	0.106
	(0.127)	(0.136)	(0.132)
Observations	3,658	3,658	3,658
R^2	0.871	0.826	0.699
Dependent Variable Mean	0.604	0.447	0.319

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variable is the share of people in the AMC in 1980 below the poverty line in terms of their per capita family income. The poverty line used in column 1, obtained from IPEA, is half the Brazilian minimum wage in September 1991, whereas the poverty lines in columns 2 and 3 (US$ 2 and US$ 1.25 a day at 2005 PPP) were taken from Ravallion (2012). The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 mean per capita family income, average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast.

Table 3.7: Income Shares in 1970 and Real Growth in Value of Firms' Capital Stocks between 1970 and 1980

	Agriculture	Commercial	Manufacturing	Services	Total
Share of 1970 AMC income held by Q5	-0.052	-1.168	-5.228**	-0.726	-0.857
	(0.596)	(0.868)	(2.052)	(1.231)	(0.538)
Share of 1970 AMC income held by Q4	0.060	-0.647	-5.063*	-0.183	0.121
	(0.751)	(1.231)	(2.599)	(1.735)	(0.730)
Share of 1970 AMC income held by Q2	0.262	0.159	-2.920	-0.504	-0.313
	(1.014)	(1.448)	(3.860)	(2.029)	(1.012)
Share of 1970 AMC income held by Q1	-2.962**	-5.480***	-8.951**	-4.900**	-3.576***
	(1.336)	(1.916)	(3.909)	(2.438)	(1.244)
Observations	3,659	3,659	3,659	3,659	3,659
R^2	0.206	0.240	0.362	0.637	0.109
P-value (Q5+Q1=0)	0.093	0.011	0.010	0.099	0.007

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variables are the 1970-1980 growth rate in the value of the AMC's private sector capital stocks for each productive sector, calculated by IPEA from the 1970 and 1980 economic censuses. The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 capital stocks in all sectors, mean per capita family income (in ln), average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast.

Table 3.8: Income Shares in 1970 and Educational Attainment in 1980

	Average years of education	% of people by years of education		
		$<$ 4 years	$\geq$ 4 and $<$ 8 years	$\geq$ 8 years
Share of 1970 AMC income held by Q5	0.323	-0.002	-0.024	0.026*
	(0.225)	(0.037)	(0.035)	(0.014)
Share of 1970 AMC income held by Q4	0.390	-0.031	0.001	0.031*
	(0.311)	(0.052)	(0.049)	(0.018)
Share of 1970 AMC income held by Q2	0.388	-0.055	0.062	-0.006
	(0.373)	(0.064)	(0.059)	(0.023)
Share of 1970 AMC income held by Q1	-1.091**	0.177**	-0.148**	-0.029
	(0.491)	(0.079)	(0.075)	(0.029)
Observations	3,659	3,659	3,659	3,659
R^2	0.935	0.925	0.897	0.886
Dependent Variable Mean	2.073	0.742	0.196	0.062

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. All dependent variables are calculated in 1980 for individuals 25 years and older. The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 capital stocks in all sectors, mean per capita family income (in ln), average schooling attainment, % of people according to educational attainment groups, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast.

Table 3.9: Income Shares in 1970, Population Growth and Migration from 1970 to 1980

	Population Growth	Immigration Rate	Emigration Rate	Fertility Rate	Mortality Rate
Share of 1970 AMC income held by Q5	-0.397	-0.167	0.240***	-0.033	-0.043
	(0.383)	(0.317)	(0.053)	(0.051)	(0.107)
Share of 1970 AMC income held by Q4	-0.004	0.211	0.142*	-0.036	0.037
	(0.427)	(0.356)	(0.073)	(0.064)	(0.135)
Share of 1970 AMC income held by Q2	0.787	0.616	0.032	0.103	-0.101
	(0.661)	(0.466)	(0.100)	(0.105)	(0.223)
Share of 1970 AMC income held by Q1	-1.494**	-0.627	0.232**	-0.150	0.483**
	(0.643)	(0.506)	(0.113)	(0.105)	(0.210)
Observations	3,659	3,659	3,659	3,659	3,659
R^2	0.210	0.194	0.524	0.596	0.276
Dependent Variable Mean	0.137	0.258	0.190	0.242	0.174

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variable in column 1 is the AMC's population growth rate in 1970-1980, and the dependent variable in column 2 is the immigration rate between 1970 and 1980, i.e., the ratio between the number of people living in the AMC in 1980 who were not living there in 1970 (or who belong to a family in which the head was not living there in 1970 if aged less than 10) and the AMC's population in 1970. The dependent variable in column 3 is the AMC's emigration rate in 1970-1980, calculated as the ratio between the number of people who reported the AMC as their previous residence but were not living there in 1980, and the AMC's population in 1970. The dependent variable in column 4 is the AMC's fertility rate in 1970-1980, computed as the ratio between the number of children less than 10 year old living in the AMC in 1980 whose parents are non-immigrants and the AMC population in 1970. The dependent variable in column 5, which we refer to as mortality rate, is the ratio between the change in population between 1970 and 1980 not accounted for by fertility and migration and the 1970 population; this is a residual category, including not only people who passed away in 1970-1980, but also individuals who emigrated from the AMC but did not report their municipality of origin. The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 mean per capita family income (in ln), average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast.

3.9 Appendix Tables

Table 3.A1: Income Inequality in 1970 and Subsequent Economic Growth (Controlling for Sectoral Labor Shares)

	1970-2000 growth		Ln (Income)		
	Ln difference		1980	1991	2000
Gini coefficient	0.333***		0.291***	0.335***	0.333***
	(0.065)		(0.080)	(0.078)	(0.065)
Gini approximation		0.395***			
		(0.071)			
Ln(1970 income)	-0.789***	-0.788***	0.359***	0.264***	0.211***
	(0.017)	(0.017)	(0.024)	(0.022)	(0.017)
Observations	3,659	3,659	3,659	3,659	3,659
R^2	0.682	0.682	0.860	0.852	0.880

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variable in columns 1 and 2 is the 1970-2000 growth in the mean per capita family income, and the dependent variable in columns 3, 4 and 5 is the mean per capita family income in 1980, 1990 and 2000 (in ln). The 1970 Gini approximation based on quintile shares is calculated using the formula in (3.2), and Ln (1970 income) is the mean per capita family income in 1970 (in ln). All regressions include state fixed effects and control for average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, whether the AMC is located on the coast, and the share of occupied individuals working in the 16 economic sectors defined in the census.

Table 3.A2: Income Shares in 1970 and Subsequent Economic Growth (Controlling for Sectoral Labor Shares)

	1970-2000 growth	Ln (Income)		
	Ln difference	1980	1991	2000
Share of 1970 AMC income held by Q5	0.053	0.167	-0.026	0.053
	(0.175)	(0.232)	(0.203)	(0.175)
Share of 1970 AMC income held by Q4	0.097	0.118	-0.106	0.097
	(0.228)	(0.332)	(0.254)	(0.228)
Share of 1970 AMC income held by Q2	-0.347	0.345	-0.132	-0.347
	(0.281)	(0.336)	(0.324)	(0.281)
Share of 1970 AMC income held by Q1	-1.244***	-1.378***	-1.622***	-1.244***
	(0.385)	(0.483)	(0.463)	(0.385)
Observations	3,659	3,659	3,659	3,659
R^2	0.683	0.861	0.853	0.881
P-value (Q4+Q2=0)	0.547	0.385	0.616	0.547
P-value (Q5+Q1=0)	0.023	0.073	0.009	0.023

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variable in column 1 is the 1970-2000 growth in the mean per capita family income, and the dependent variable in columns 3, 4 and 5 is the mean per capita family income in 1980, 1990 and 2000 (in ln). The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 mean per capita family income, average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, whether the AMC is located on the coast, and the share of occupied individuals working in the 16 economic sectors defined in the census.

Table 3.A3: Income Shares in 1970 and Real Growth in Value of Firms' Capital Stocks between 1970 and 1980 (Controlling for Sectoral Labor Shares)

	Agriculture	Commercial	Manufacturing	Services	Total
Share of 1970 AMC income held by Q5	0.055	-0.952	-3.572*	-0.400	-0.649
	(0.562)	(0.862)	(1.992)	(1.206)	(0.532)
Share of 1970 AMC income held by Q4	0.243	-0.709	-4.404*	-0.343	0.209
	(0.734)	(1.207)	(2.487)	(1.682)	(0.728)
Share of 1970 AMC income held by Q2	0.234	0.521	-2.594	0.205	-0.194
	(0.996)	(1.400)	(3.809)	(1.966)	(1.007)
Share of 1970 AMC income held by Q1	-3.026**	-4.433**	-5.183	-2.742	-3.188**
	(1.295)	(1.917)	(3.913)	(2.447)	(1.256)
Observations	3,659	3,659	3,659	3,659	3,659
R^2	0.230	0.271	0.387	0.651	0.119
P-value (Q5+Q1=0)	0.081	0.039	0.109	0.355	0.020

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variables are the 1970-1980 growth rate in the value of the AMC's private sector capital stocks for each productive sector, calculated by IPEA from the 1970 and 1980 economic censuses. The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 capital stocks in all sectors, mean per capita family income (in ln), average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, whether the AMC is located on the coast, and the share of occupied individuals working in the 16 economic sectors defined in the census.

Table 3.A4: Income Shares in 1970 and Educational Attainment in 1980 (Controlling for Sectoral Labor Shares)

	Average years of education	% of people by years of education		
		$<$ 4 years	$\geq$ 4 and $<$ 8 years	$\geq$ 8 years
Share of 1970 AMC income held by Q5	0.270	0.002	-0.027	0.025*
	(0.230)	(0.037)	(0.035)	(0.013)
Share of 1970 AMC income held by Q4	0.336	-0.026	0.003	0.024
	(0.319)	(0.052)	(0.048)	(0.018)
Share of 1970 AMC income held by Q2	0.533	-0.072	0.061	0.011
	(0.369)	(0.063)	(0.057)	(0.021)
Share of 1970 AMC income held by Q1	-0.947*	0.163**	-0.162**	-0.001
	(0.492)	(0.080)	(0.076)	(0.027)
Observations	3,659	3,659	3,659	3,659
R^2	0.937	0.927	0.899	0.897
Dependent Variable Mean	2.073	0.742	0.196	0.062

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. All dependent variables are calculated in 1980 for individuals 25 years and older. The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 capital stocks in all sectors, mean per capita family income (in ln), average schooling attainment, % of people according to educational attainment groups, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, whether the AMC is located on the coast, and the share of occupied individuals working in the 16 economic sectors defined in the census.

Table 3.A5: Income Shares in 1970 and Subsequent Economic Growth (Imputing Top-Coded Incomes)

	1970-2000 growth	Ln (Income)		
	Ln difference	1980	1991	2000
Share of 1970 AMC income held by Q5	0.191	0.123	0.033	0.191
	(0.170)	(0.227)	(0.198)	(0.170)
Share of 1970 AMC income held by Q4	0.221	0.127	-0.021	0.221
	(0.229)	(0.330)	(0.256)	(0.229)
Share of 1970 AMC income held by Q2	-0.377	0.324	-0.148	-0.377
	(0.278)	(0.337)	(0.321)	(0.278)
Share of 1970 AMC income held by Q1	-1.117***	-1.494***	-1.651***	-1.117***
	(0.382)	(0.471)	(0.454)	(0.382)
Observations	3,659	3,659	3,659	3,659
R^2	0.677	0.858	0.849	0.877
P-value (Q4+Q2=0)	0.705	0.400	0.722	0.705
P-value (Q5+Q1=0)	0.072	0.037	0.008	0.072

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. The dependent variable in column 1 is the 1970-2000 growth in the mean per capita family income, and the dependent variable in columns 3, 4 and 5 is the mean per capita family income in 1980, 1990 and 2000 (in ln). The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 mean per capita family income, average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast. Top-coded incomes in 1970 are multiplied by a factor of 2.15 so that individual incomes in the top 20% follow a Pareto distribution.

Table 3.A6: Income Shares in 1970 and Subsequent Economic Growth (Including Collective Households and Non-Family Members)

	1970-2000 growth	Ln (Income)		
	Ln difference	1980	1991	2000
Share of 1970 AMC income held by Q5	0.314**	0.438*	0.267	0.314**
	(0.142)	(0.246)	(0.208)	(0.142)
Share of 1970 AMC income held by Q4	0.419**	0.547	0.326	0.419**
	(0.196)	(0.339)	(0.275)	(0.196)
Share of 1970 AMC income held by Q2	-0.307	0.583	0.014	-0.307
	(0.272)	(0.377)	(0.337)	(0.272)
Share of 1970 AMC income held by Q1	-0.942***	-1.119**	-1.136**	-0.942***
	(0.356)	(0.505)	(0.472)	(0.356)
Observations	3,659	3,659	3,659	3,659
R^2	0.672	0.859	0.849	0.878
P-value (Q5+Q1=0)	0.170	0.338	0.174	0.170

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses. The unit of observation is an AMC. For the calculation of AMC per capita family income in 1970 we do not exclude individuals living in collective dwellings and those living in private dwellings which are unrelated to the family head. The dependent variable in column 1 is the 1970-2000 growth in the mean per capita family income, and the dependent variable in columns 3, 4 and 5 is the mean per capita family income in 1980, 1990 and 2000 (in ln). The explanatory variables of interest are the shares of 1970 AMC income held by each of the quintiles, with the third quintile being the omitted category. All regressions include state fixed effects and control for 1970 mean per capita family income, average schooling attainment, literacy rate, population, % of urban population, life expectancy, latitude, longitude, distance from state and federal capital, and whether the AMC is located on the coast.

Bibliography

Aghion, Philippe and Patrick Bolton, "A Theory of Trickle-Down Growth and Development," *The Review of Economic Studies*, 1997, *64* (2), 151–172.

Alesina, Alberto, "Credibility and Policy Convergence in a Two-Party System with Rational Voters," *American Economic Review*, 1988, *78* (4), 796–805.

Alesina, Alberto and Dani Rodrik, "Distributive Politics and Economic Growth," *The Quarterly Journal of Economics*, 1994, pp. 465–490.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Araujo, M Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo, and Norbert Schady, "Teacher Quality and Learning Outcomes in Kindergarten," *Quarterly Journal of Economics*, 2016, *131* (3), 1415–1453.

Autor, David H. and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 2013, *103* (5), 1553–1597.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, "Trends in US Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 2008, *90* (2), 300–323.

Azmat, Ghazala and Nagore Iriberri, "The Importance of Relative Performance Feedback Information: Evidence From a Natural Experiment using High School Students," *Journal of Public Economics*, 2010, *94* (7), 435–452.

Baird, Jo-Anne, "What's in a Name? Experiments with Blind Marking in A-Level Examinations," *Educational Research*, 1998, *40* (2), 191–202.

Baker, George, "Distortion and Risk in Optimal Incentive Contracts," *Journal of Human Resources*, 2002, pp. 728–751.

Baker, George P, "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, 1992, *100* (3), 598–614.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Preferences and the Response To Incentives: Evidence From Personnel Data," *Quarterly Journal of Economics*, 2005, *120* (3), 917–962.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Team Incentives: Evidence From a Firm Level Experiment," *Journal of the European Economic Association*, 2013, *11* (5), 1079–1114.

Bandiera, Oriana, Iwan Barankay, Imran Rasul et al., "Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment," *Quarterly Journal of Economics*, 2007, *122* (2), 729–773.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton, "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India," 2016. NBER Working Paper 22746.

Banerjee, Abhijit V. and Andrew F. Newman, "Occupational Choice and the Process of Development," *Journal of Political Economy*, 1993, *101* (2), 274–298.

Barrera-Osorio, Felipe and Dhushyanth Raju, "Teacher Performance Pay: Experimental Evidence from Pakistan," *Journal of Public Economics*, 2017, *148*, 75–91.

Barro, Robert J., "Inequality and Growth in a Panel of Countries," *Journal of Economic Growth*, 2000, *5* (1), 5–32.

Bechtel, Michael M., Dominik Hangartner, and Lukas Schmid, "Compulsory Voting, Habit Formation, and Political Participation," 2016. Working Paper.

Bechtel, Michael M., Dominik Hangartner, and Lukas Schmid, "Does Compulsory Voting Increase Support for Leftist Policy?," *American Journal of Political Science*, 2016, *60* (3), 752–767.

Behrman, Jere R, Susan W Parker, Petra E Todd, and Kenneth I Wolpin, "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools," *Journal of Political Economy*, 2015, *123* (2), 325–364.

Benabou, Roland, "Inequality and Growth," in "NBER Macroeconomics Annual 1996, Volume 11," MIT Press, 1996, pp. 11–92.

Benjamin, Dwayne, Loren Brandt, and John Giles, "Did Higher Inequality Impede Growth in Rural China?," *Economic Journal*, 2011, *121* (557), 1281–1309.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Besley, Timothy and Stephen Coate, "An Economic Model of Representative Democracy," *Quarterly Journal of Economics*, 1997, *112* (1), 85–114.

Bischof, Günter and Fritz Plasser, *The changing Austrian voter*, Vol. 16, Transaction Publishers, 2008.

Blöndal, Jón R. and Daniel Bergvall, "Budgeting in Austria," *OECD Journal on Budgeting*, 2007, *7*, 1–37.

Börgers, Tilman, "Costly Voting," *American Economic Review*, 2004, *94* (1), 57–66.

Botelho, Fernando, Ricardo A Madeira, and Marcos A Rangel, "Racial Discrimination in Grading: Evidence from Brazil," *American Economic Journal: Applied Economics*, 2015, *7* (4), 37–52.

Bourguignon, Francois, "Pareto Superiority of Unegalitarian Equilibria in Stiglitz' Model of Wealth Distribution with Convex Saving Function," *Econometrica*, 1981, pp. 1469–1475.

Brady, Henry and John McNulty, "Turning Out the Vote: The Costs of Finding and Getting to the Polling Place," *American Political Science Review*, 2011, *5* (1), 1–20.

Bruns, Barbara and Javier Luque, *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean*, World Bank Publications, 2015.

Bruns, Barbara, Deon Filmer, and Harry Anthony Patrinos, *Making Schools Work: New Evidence on Accountability Reforms*, World Bank Publications, 2011.

Burgess, Simon and Ellen Greaves, "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities," *Journal of Labor Economics*, 2013, *31* (3), 535–576.

Calsamiglia, Caterina and Annalisa Loviglio, "Maturity and School Outcomes in an Inflexible System: Evidence from Catalonia," 2016. Working Paper.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, May 2008, *90* (3), 414–427.

Campante, Filipe R. and Francisco H.G. Ferreira, "Inefficient Lobbying, Populism and Oligarchy," *Journal of Public Economics*, 2007, *91* (5), 993–1021.

Cascio, Elizabeth U. and Ebonya Washington, "Valuing the Vote: The Redistribution of Voting Rights and State Funds Following the Voting Rights Act of 1965," *The Quarterly Journal of Economics*, 2014, *129* (1), 376–433.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F Halsey Rogers, "Missing in Action :Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, 2006, *20* (1), 91–116.

Chetty, Raj, John N Friedman, and Jonah E Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 2014, *104* (9), 2633–2679.

Chevalier, Judith and Glenn Ellison, "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, 1997, *105* (6), 1167–1200.

Chong, Alberto and Mauricio Olivera, "Does Compulsory Voting Help Equalize Incomes?," *Economics and Politics*, 2008, *20* (3), 391–415.

Chong, Alberto, Isabelle Cohen, Erica Field, Eduardo Nakasone, and Maximo Torero, "Iron Deficiency and Schooling Attainment in Peru," *American Economic Journal: Applied Economics*, 2016.

Contreras, Dante and Tomás Rau, "Tournament Incentives for Teachers: Evidence from a Scaled-Up Intervention in Chile," *Economic Development and Cultural Change*, 2012, *61* (1), 219–246.

Dal Bó, Ernesto, Frederico Finan, and Martín A Rossi, "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *Quarterly Journal of Economics*, 2013, *128* (3), 1169–1218.

De Leon, Fernanda Leite Lopez and Renatai Rizzi, "A Test for the Rational Ignorance Hypothesis: Evidence from a Natural Experiment in Brazil," *American Economic Journal: Economic Policy*, 2014, *6* (4), 380–398.

Deaton, Angus and Nancy Cartwright, "Understanding and Misunderstanding Randomized Controlled Trials," 2016. National Bureau of Economic Research Working Paper w22595.

Deininger, Klaus and Lyn Squire, "New Ways of Looking at Old Issues: Inequality and Growth," *Journal of Development Economics*, 1998, *57* (2), 259–287.

Deserranno, Erika, "Financia Incentives as Signals: Experimental Evidence From the Recruitment of Health Workers," 2016. Working Paper.

Easterly, William, "Inequality Does Cause Underdevelopment: Insights from a New Instrument," *Journal of Development Economics*, 2007, *84* (2), 755–776.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya, "Media and Political Persuasion: Evidence from Russia," *American Economic Review*, 2010, *111* (7), 3253–3285.

Esteban, Joan and Debraj Ray, "Wealth Constraints, Lobbying and the Efficiency of Public Allocation," *European Economic Review*, 2000, *44* (4), 694–705.

Farber, Henry, "Increasing Voter Turnout: Is Democracy Day the Answer?," 2009. Princeton University Working Paper 546.

Ferreira, Francisco H. G., Christoph Lakner, Maria Ana Lugo, and Berk Özler, "Inequality of Opportunity and Economic Growth: A Cross-Country Analysis," 2014.

Ferwerda, Jeremy, "Electoral Consequences of Declining Participation: A Natural Experiment in Austria," *Electoral Studies*, 2014, *35*, 242–252.

Figlio, David N, "Testing, crime and punishment," *Journal of Public Economics*, 2006, *90* (4), 837–851.

Figlio, David N and Joshua Winicki, "Food for thought: the effects of school accountability plans on school nutrition," *Journal of Public Economics*, 2005, *89* (2), 381–394.

Forbes, Kristin J., "A Reassessment of the Relationship between Inequality and Growth," *American Economic Review*, 2000, pp. 869–887.

Fowler, Anthony, "Electoral and Political Consequences of Voter Turnout: Evidence from Compulsory Voting in Australia," *Quarterly Journal of Political Science*, 2013, *8* (2), 159–182.

Fraga, Bernard and Eitan Hersh, "Voting Costs and Voter Turnout in Competitive Elections," *Quarterly Journal of Political Science*, 2010, *5*, 339–356.

Fryer, Roland G, "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools," *Journal of Labor Economics*, 2013, *31* (2), 373–407.

Fuentes, Andrew, Eckhard Wurzel, and A. Wörgötter, "Reforming Federal Fiscal Relations in Austria," 2006. OECD Economics Department Working Papers, No. 474, OECD Publishing.

Fujiwara, Thomas, "Voting Technology, Political Responsiveness, and Infant Health: Evidence from Brazil," *Econometrica*, 2015, *83* (2), 423–464.

Funk, Patricia, "Is there an Expressive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines," *American Law and Economics Review*, 2007, *9* (2), 135–159.

Gallego, Aina, "Unequal Political Participation in Europe," *International Journal of Sociology*, 2007, *37* (4), 10–25.

Galor, Oded and Joseph Zeira, "Income Distribution and Macroeconomics," *Review of Economic Studies*, 1993, *60* (1), 35–52.

Galor, Oded and Omer Moav, "From Physical to Human Capital Accumulation: Inequality and the Process of Development," *The Review of Economic Studies*, 2004, *71* (4), 1001–1026.

Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti, "Internet and Politics: Evidence from UK Local Elections and Local Government Policies," 2014.

Gentzkow, Matthew, "Television and Voter Turnout," *Quarterly Journal of Economics*, 2006, *121* (3), 931–972.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson, "The Effect of Newspaper Entry and Exit on Electoral Politics," *American Economic Review*, 2011, *101* (7), 931–972.

Gerber, Alan S and Donald P Green, *Field experiments: Design, analysis, and interpretation*, WW Norton, 2012.

Geys, Benny, "Explaining voter turnout: A review of aggregate-level research," *Electoral Studies*, 2006, *25*, 637–663.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics*, 2010, *2* (3), 205–227.

Gneezy, Uri and Aldo Rustichini, "A Fine is a Price," *Journal of Legal Studies*, 2000, *29* (1).

Gneezy, Uri and Aldo Rustichini, "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 2000, *115* (3), 791–810.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel, "When and Why Incentives (Don't) Work to Modify Behavior," *Journal of Economic Perspectives*, 2011, *25* (4), 191–209.

Godefroy, Rafael and Emeric Henry, "Voter turnout, politicians' quality and public expenditures," 2015. Mimeo.

Gomez, Brad, Thomas Hansford, and George Krause, "The Republicans Should Pray for Rain: Weather, Turnout, and Voting in U.S. Presidential Elections," *Journal of Politics*, 2007, *69*, 649–663.

Goodman, Sarena F and Lesley J Turner, "The Design of Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program," *Journal of Labor Economics*, 2013, *31* (2), 409–420.

Gradstein, Mark and Branko Milanovic, "Does Liberté = Egalité? A Survey of the Empirical Links between Democracy and Inequality with Some Evidence on the Transition Economies," *Journal of Economic Surveys*, 2004, *18* (4), 515–537.

Groves, Theodore, Yongmiao Hong, John McMillan, and Barry Naughton, "Autonomy and Incentives in Chinese State Enterprises," *Quarterly Journal of Economics*, 1994, *109* (1), 183–209.

Hagopian, Frances, *Traditional Politics and Regime Change in Brazil*, Cambridge University Press, 1996.

Haller, Max, Kurt Holm, and R. Oldenbourg Verlag, *Werthaltungen und Lebensformen in Österreich. Ergebnisse des Sozialen Survey 1986*, Munich: R. Oldenbourg Verlag, 1987.

Haller, Max, Wolfgang Schulz, and Alfred Grausgruber, *Österreich zur Jahrhundertwende. Gesellschaftliche Werthaltungen und Lebensqualität 1986–2004*, Wiesbaden: Verlag für Sozialwissenschaften, 2005.

Hanna, Rema N and Leigh L Linden, “Discrimination in Grading,” *American Economic Journal: Economic Policy*, 2012, *4* (4), 146–168.

Hansford, Thomas and Brad Gomez, “Estimating the Electoral Effects of Voter Turnout,” *American Political Science Review*, 2010, *104* (2), 268–288.

Hanushek, Eric A and Steven G Rivkin, “Generalizations about Using Value-Added Measures of Teacher Quality,” *American Economic Review*, 2010, *100* (2), 267–271.

Herrera, Helios, Massimo Morelli, and Thomas Palfrey, “Turnout and power sharing,” *The Economic Journal*, 2014, *124* (574), F131–F162.

Hinnerich, Björn Tyrefors and Per Pettersson-Lidbom, “Democracy, Redistribution, and Political Participation: Evidence From Sweden 1919-1938,” *Econometrica*, 2014, *82* (3), 961–993.

Hirczy, Wolfgang, “The Impact of Mandatory Voting Laws on Turnout: A Quasi-Experimental Approach,” *Electoral Studies*, 1994, *13* (1), 64–76.

Hodler, Roland, Simon Luechinger, and Alois Stutzer, “The Effects of Voting Costs on the Democratic Process and Public Finances,” *American Economic Journal: Economic Policy*, 2015, *7* (1), 141–171.

Holmstrom, Bengt, “Moral Hazard in Teams,” *Bell Journal of Economics*, 1982, pp. 324–340.

Holmstrom, Bengt and Paul Milgrom, “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design,” *Journal of Law, Economics, & Organization*, 1991, *7*, 24–52.

Husted, Thomas A. and Lawrence W. Kenny, “The Effect of the Expansion of the Voting Franchise on the Size of Government,” *Journal of Political Economy*, 1997, *105*, 54–82.

Imberman, Scott A and Michael F Lovenheim, “Incentive Strength and Teacher Productivity: Evidence from a Group-Based Teacher Incentive Pay System,” *Review of Economics and Statistics*, 2015, *97* (2), 364–386.

IMF, “Austria: Selected Issues,” 2008. International Monetary Fund Country Report No. 08/189.

Jacob, Brian A and Steven D Levitt, “Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating,” *Quarterly Journal of Economics*, 2003, *118* (3), 843–877.

Kandel, Eugene and Edward P Lazear, “Peer Pressure and Partnerships,” *Journal of Political Economy*, 1992, *100* (4), 801–817.

Kandori, Michihiro, “Social Norms and Community Enforcement,” *Review of Economic Studies*, 1992, *59* (1), 63–80.

Katz, Lawrence F. and Kevin M. Murphy, “Changes in Relative Wages, 1963–1987: Supply and Demand Factors,” *Quarterly Journal of Economics*, 1992, *107* (1), 35–78.

Kerwin, Jason T and Rebecca Thornton, “Making the Grade: Understanding What Works for Teaching Literacy in Rural Uganda,” 2015.

Knack, Steve, “Does Rain Help Republicans? Theory and Evidence on Turnout and Vote,” *Public Choice*, 1994, *79*, 187–209.

Koretz, Daniel M, “Limitations in the Use of Achievement Tests as Measures of Educators’ Productivity,” *Journal of Human Resources*, 2002, *37* (4), 752–777.

Krasa, Stefan and Matthias K. Polborn, “Is mandatory voting better than voluntary voting?,” *Games and Economic Behavior*, 2009, *66* (1), 275–291.

Krishna, Vijay and John Morgan, “Overcoming Ideological Bias in Elections,” *Journal of Political Economy*, 2011, *119* (2), 183–211.

Kritzinger, Sylvia, Eva Zeglovits, Michael S Lewis-Beck, and Richard Nadeau, *The Austrian Voter*, Vandenhoeck & Ruprecht, 2013.

Lavy, Victor, “Evaluating the Effect of Teachers’ Group Performance Incentives on Pupil Achievement,” *Journal of Political Economy*, 2002, *110* (6), 1286–1317.

Lavy, Victor, "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence From a Natural Experiment," *Journal of Public Economics*, 2008, *92* (10), 2083–2105.

Lavy, Victor, "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics," *American Economic Review*, 2009, *99* (5), 1979–2021.

Lavy, Victor, "Teachers' Pay for Performance in the Long-Run: Effects on Students' Educational and Labor Market Outcomes in Adulthood," 2015. NBER Working Paper 20983.

Lazear, Edward P and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 1981, *89* (5), 841–864.

Lee, David S., Enrico Moretti, and Matthew J. Butler, "Do Voters Affect Or Elect Policies? Evidence from the U. S. House," *Quarterly Journal of Economics*, 2004, *119* (3), 807–859.

Lehner, Gerhard, *Die Bundesländer im Finanzausgleich*, WIFO: Austrian Institute for Economic Research, 1997.

León, Gianmarco, "Turnout, Political Preferences and Information: Experimental Evidence from Peru," *Journal of Development Economics*, 2017, *127*, 56–71.

Li, Hongyi, Lyn Squire, and Heng-Fu Zou, "Explaining International and Intertemporal Variations in Income Inequality," *The Economic Journal*, 1998, *108* (446), 26–43.

Lijphart, Arend, "Unequal Participation: Democracy's Unresolved Dilemma," *American Political Science Review*, 1997, *91* (1), 1–14.

Linz, Juan, Alfred Stephan, and Yogendra Yadav, *Democracy and Diversity*, New Delhi: Oxford University Press, 2007.

Mahler, Vincent A, David K Jesuit, and Piotr R Paradowski, "Electoral Turnout and State Redistribution A Cross-National Study of Fourteen Developed Countries," *Political Research Quarterly*, 2014, *67* (2), 361–373.

Martinez, Michael D and Jeff Gill, "The effects of turnout on partisan outcomes in US presidential elections 1960–2000," *Journal of Politics*, 2005, *67* (4), 1248–1274.

Meltzer, Allan H. and Scott F. Richard, "A Rational Theory of the Size of Government," *Journal of Political Economy*, 1981, *89* (5), 914–927.

Miller, Grant, "Women's Sufrage, Political Responsiveness, and Child Survival in American History," *The Quarterly Journal of Economics*, 2008, *123* (3), 1287–1327.

Mizala, Alejandra and Ben Ross Schneider, "Negotiating Education Reform: Teacher Evaluations and Incentives in Chile (1990–2010)," *Governance*, 2014, *27* (1), 87–109.

Mizala, Alejandra and Hugo Ñopo, "Measuring the Relative Pay of School Teachers in Latin America 1997–2007," *International Journal of Educational Development*, 2016, *47*, 20–32.

Muralidharan, Karthik and Venkatesh Sundararaman, "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India," *Economic Journal*, 2010, *120* (546), F187–F203.

Muralidharan, Karthik and Venkatesh Sundararaman, "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39–77.

Naidu, Suresh, "Suffrage, Schooling, and Sorting in the Post-Bellum U.S. South," 2012. NBER Working Paper18129.

Neal, Derek, "The Design of Performance Pay in Education," *Handbook of the Economics of Education*, 2011, *4*, 495–550.

Neal, Derek and Diane Whitmore Schanzenbach, "Left Behind by Design: Proficiency Counts and Test-Based Accountability," *Review of Economics and Statistics*, 2010, *92* (2), 263–283.

Newstead, Stephen E and Ian Dennis, “Blind Marking and Sex Bias in Student Assessment,” *Assessment and Evaluation in higher Education*, 1990, *15* (2), 132–139.

OECD, “Managing across Levels of Government: Austria,” 1997, pp. 93–106.

Osborne, Martin J. and Al Silvinski, “A Model of Political Competition with Citizen-Candidates,” *The Quarterly Journal of Economics*, 1996, *111* (01), 65–96.

Oster, Emily, “Unobservable Selection and Coefficient Stability: Theory and Evidence,” *Journal of Business Economics and Statistics*, Forthcoming.

Oyer, Paul, “Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality,” *Quarterly Journal of Economics*, 1998, *113* (1), 149–185.

Panizza, Ugo, “Income Inequality and Economic Growth: Evidence from American Data,” *Journal of Economic Growth*, 2002, *7* (1), 25–41.

Partridge, Mark D., “Is Inequality Harmful for Growth? Comment,” *The American Economic Review*, 1997, *87* (5), 1019–1032.

Perotti, Roberto, “Growth, Income Distribution, and Democracy: What the Data say,” *Journal of Economic growth*, 1996, *1* (2), 149–187.

Persson, Torsten and Guido Tabellini, “Is Inequality Harmful for Growth?,” *The American Economic Review*, 1994, pp. 600–621.

Persson, Torsten and Guido Tabellini, *Political Economics: Explaining Economic Policy*, Cambridge MA: MIT Press, 2000.

Prendergast, Canice, “The Provision of Incentives in Firms,” *Journal of Economic Literature*, 1999, *37* (1), 7–63.

Ravallion, Martin, “Why Don’t We See Poverty Convergence?,” *The American Economic Review*, 2012, *102* (1), 504–523.

Ravallion, Martin, Shaohua Chen, and Prem Sangraula, "Dollar a Day Revisited," *World Bank Economic Review*, 2009, p. lhp007.

Reis, Eustáquio, Kepler Magalhães, Márcia Pimentel, and Mérida Medina, "Estoque de Capital Privado nos Município Brasileiros, 1970-85," *Rio de Janeiro, IPEA*, 2005.

Rockoff, Jonah E, "The Impact of Individual Teachers on Student Achievement: Evidence From Panel Data," *American Economic Review*, 2004, *94* (2), 247–252.

Rodriguez, F. C, "Does Distributional Skewness Lead to Redistribution? Evidence from the United States," *Economics and Politics*, 1999, *11*, 171–199.

Shineman, Victoria, "Isolating the Effect of Compulsory Voting Laws on Political Sophistication: Exploiting Intra-National Variation in Mandatory Voting Laws Between the Austrian Provinces," 2014.

Springer, Matthew G, Laura Hamilton, Daniel F McCaffrey, Dale Ballou, Vi-Nhuan Le, Matthew Pepper, JR Lockwood, and Brian M Stecher, "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching.," *National Center on Performance Incentives*, 2010.

Stromberg, David, "Radio's Impact on Public Spending," *Quarterly Journal of Economics*, 2004, *119* (1), 589–621.

Timpone, Richard J., "Structure, Behavior, and Voter Turnout in the United States," *The American Political Science Review*, 1998, *92* (1), 145–158.

Timpone, Richard J., "Demokratische Beteiligung und Staatsausgaben: Die Auswirkungen des Frauenstimmrechts," *Swiss Journal of Economics and Statistics*, 2005, *141* (4), 617–650.

Tran, Anh and Richard Zeckhauser, "Rank as an Inherent Incentive: Evidence From a Field Experiment," *Journal of Public Economics*, 2012, *96* (9), 645–650.

Van Ewijk, Reyn, “Same Work, Lower Grade? Student Ethnicity and Teachersâ Subjective Assessments,” *Economics of Education Review*, 2011, *30* (5), 1045–1058.

Voitchovsky, Sarah, “Does the Profile of Income Inequality Matter for Economic Growth?,” *Journal of Economic Growth*, 2005, *10* (3), 273–296.

Washington, Ebonya, “How Black Candidates Affect Voter Turnout,” *Quarterly Journal of Economics*, 2006, *121* (3), 979–998.