# THÈSE

En vue de l'obtention du

# DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l' Université Toulouse 1 Capitole
Discipline : Sciences Economiques

---

Présentée et soutenue par

**Margaret Alice LEIGHTON**
Le 6 juillet 2015

Titre :

## Essays on the Economics of Education

---

Ecole doctorale : Toulouse School of Economics

Unité de recherche : ARQADE – TSE

Directeur de Thèse : Stéphane Straub

JURY

Liliane BONNAL, professeur, Université de Poitiers
Sylvain CHABE-FERRET, Ingénieur en Chef des Ponts, des Eaux et des Forêts, Inra (Lerna) et Toulouse School of Economics
Sergio P. FIRPO, professeur associé, São Paulo School of Economics
Julien GRENET, chargé de recherche au CNRS, professeur associé à PSE-École d'Économie de Paris
Christian HELLWIG, professeur, Université Toulouse 1 et Toulouse School of Economics
Stéphane STRAUB, professeur, Université Toulouse 1 et Toulouse School of Economics

# Essays on the Economics of Education: Introduction

The existence of causal returns of education on wages is both deeply intuitive and supported by extensive evidence. Decades of research have been devoted to estimating the magnitude of these returns. Despite this large body of research, relatively little is known about *how* education increases wages. Our intuition provides some preliminary channels: a man who can read can do a great many things which are impossible for an illiterate man; a woman who cannot add or subtract will face challenges which a numerate woman handles with ease. Similarly, a surgeon who has mastered two different surgeries is more versatile than her colleague who has mastered only one; a teacher who has expertise in chemistry and biology can teach both courses, and can exploit complementaries between the two.

While such intuitions seemingly help justify the large returns econometricians have estimated for each additional year of schooling, they are at best partial descriptions of the human capital accumulated at school. The vast majority of formal education is not spent learning how to do things which are done at work. The surgeon spent 12 or 16 years at school before starting to learn how to do surgery. The bio-chem teacher spent more time in classes unrelated to biology and chemistry than he did learning the things he teaches. If asked to identify which of their daily work tasks were acquired at school, many tertiary-educated individuals would be hard pressed to come up with anything beyond the "three Rs": reading, writing and arithmetic. These skills are rarely explicitly taught beyond primary school. What are we to make, then, of the returns to education? What are students actually learning?

The first chapter of this thesis considers this question directly. In the paper, we estimate the importance of two aspects of human capital accumulation in US tertiary education: the acquisition of job-related skills, and the student's discovery of his relative abilities across disciplines. Specifically, we measure whether additional years of multi-disciplinary education help students make a better choice of specialization, and at what cost in foregone specialized human capital. We document that, in the cross section, students who choose their major later are more likely to change fields on the labor market. We then build and estimate a dynamic model of college education which captures the tradeoff between discovering comparative advantage and acquiring occupation-specific skills. Estimates suggest that delaying specialization is informative, although noisy. Working in the field of comparative advantage accounts for up to 20% of a well-matched worker's earnings. While education is transferable across fields with only a 10% penalty, workers who wish to change fields incur a large, one-time cost.

The second chapter considers the impacts of the education path on outcomes in primary school: in particular, how does automatically promoting young children from one grade level to the next affect retention and grade progression? Exploiting variation in grade repetition practices in Brazil, we study the effect of automatic promotion cycles on grade attainment and academic persistence of primary school children. The dynamic policy environment allows us to estimate the impact of the policy when applied at different times during schooling, both in the short term and as children exposed to the policy progress through primary school. We find that automatic promotion increases grade attainment: one year of exposure to the policy is associated with 3 students out of 100 studying one grade-level above where they would be absent the policy. This effect persists over time, and cumulates with further exposure to the policy.

The third chapter moves away from students to focus on education infrastructure. In the paper we seek to answer the question of how transfers from the federal government in Brazil affect both education spending and the resources available for education at the municipal level. We find that increased transfers lead to an immediate rise in current and capital spending. These increases are focused on education and welfare expenditure in poorer municipalities, while richer municipalities expand capital spending in the transport and housing sectors. Furthermore, particularly in wealthier municipalities, increases in transfers cause a short-term increase in local tax revenues. Positive transfer shocks are associated with increases in the number of teachers and, to a lesser extent, the number of classrooms. Transfers are also associated with substantial re-allocation of resources across schools offering classes at different levels, with secondary schools and schools teaching senior primary grades expanding at the expense of junior primary schools.

These papers shed some light on the obscure process of human capital accumulation in the classroom; however, many big questions remain. While our intuition suggests that such channels must be important, linking skills learned at school with on-the-job tasks remains difficult. Without a clearer understanding of how coursework builds individual capacities, those skills which are difficult to measure may be edged out. The influence of the structure of the school system – in terms of timing and depth of specialization, tracking and incentives, among others – is equally under-explored. A student's path through school can have complex impacts not only on skills acquisition and grades, but also on important later-life outcomes, such as higher education decisions, occupation choice, and eventually wages. In light of the pressure to increase the formal content of education, and the associated growth of test-based evaluation methods in schools, better answers to these big questions are urgently needed.

# Acknowledgments

July 6, 2015

Many ducks and many bottles of wine helped make this thesis possible. As they are no longer with us, I would like to say a few words to those who are.

First and foremost, I would like to thank Stéphane Straub and Christian Hellwig: your time and support, both during my Ph. D. and through the job market, is responsible for getting me here today. I feel fortunate, as well, for having been part of ARQADE, and thank the members for creating such a warm and welcoming community. I am also grateful to those faculty and staff who have taken time for me over the last six years: your kindness and generosity, as well as your insights and feedback, have helped me grow. I benefited greatly from the semester I spent at Yale, and I thank in particular Joseph Altonji and Melissa Tartari for making it such an enriching experience. Finally, to my Mount Allison professors: thank you for challenging and inspiring me. I aspire to be more like you.

There are not enough pages in this thesis to thank all the friends who have helped me along the way. Thanks to you, I learned enough to survive the M2 and, for better or for worse, learned to love being surrounded by economists. In Toulouse and in New Haven, and on the occasional adventure, you have made these six years memorable in all the best ways. Merci aussi aux amis musiciens: les weekends de musique dans les quatres coins du sud-ouest ont alimenté les semaines de travail, et ont été parmi les plus grands plaisirs de ces dernières années.

Last but not least, I thank my family. You made everything - and anything - possible, and that is the greatest gift of all. Thank you Anna and Ted, Patrick and Jessica (and Thomas and Nicolas!), and of course, thank you Luc.

# THE MAJOR DECISION: LABOR MARKET IMPLICATIONS OF THE TIMING OF SPECIALIZATION IN COLLEGE

LUC BRIDET AND MARGARET LEIGHTON

Abstract. College students in the United States choose their major much later than their counterparts in Europe. American colleges also typically allow students to choose when they wish to make their major decision. In this paper we estimate the benefits of such a policy: specifically, whether additional years of multi-disciplinary education help students make a better choice of specialization, and at what cost in foregone specialized human capital. We first document that, in the cross section, students who choose their major later are more likely to change fields on the labor market. We then build and estimate a dynamic model of college education where the optimal timing of specialization reflects a tradeoff between discovering comparative advantage and acquiring occupation-specific skills. Multi-disciplinary education allows students to learn about their comparative advantage, while specialized education is more highly valued in occupations related to that field. Estimates suggest that delaying specialization is informative, although noisy. Working in the field of comparative advantage accounts for up to 20% of a well-matched worker's earnings. While education is transferable across fields with only a 10% penalty, workers who wish to change fields incur a large, one-time cost. We then use these estimates to compare the current college system to one which imposes specialization at college entry. In this counterfactual, the number of workers who switch fields drops from 24% to 20%; however, the share of workers who are not working in the field of their comparative advantage rises substantially, from 23% to 30%. Overall, expected earnings fall by 1.5%.

*Date*: March 31, 2015.

The results presented in this draft are preliminary. We thank our advisors, Christian Hellwig and Stéphane Straub. We have benefited from discussions with a great number of people, particularly Joseph Altonji, Melissa Tartari and Paul Rodríguez-Lesmes as well as colleagues and seminar participants at the Toulouse School of Economics, Yale, Princeton EconCon, the ENTER Jamboree, the Summer School on Socioeconomic Inequality and Mount Allison University. We are additionally grateful to Joseph Altonji and the National Center for Education Statistics staff for making it possible for us to access to the Baccalaureate and Beyond 93:03 dataset.

Affiliation: Toulouse School of Economics. Contact: margaret.leighton@tse-fr.eu, Luc.Bridet@ut-capitole.fr.

# 1. Introduction

On the first day of college at an American university, many freshmen do not know what field they will concentrate on during their undergraduate studies. Four years later, newly-minted graduates enter the world of work with a new credential and a field of specialization. Despite the dire importance students ascribe to their choice of major, relatively little research has examined the process by which this decision is reached and the implications of constraining that process.

Education systems differ widely in how and when students are allowed to select a field. In many european universities the choice of major is made prior to enrollment and is difficult to adjust thereafter. Other countries, in contrast, are much more forgiving towards the undecided. It is well accepted in the United States that college is a time of self discovery: the exploration of different fields is encouraged and sometimes mandated. In the US majors can be chosen several years into college, and adjusted even later.

The impact of constraining the timing of specialization on eventual labor market outcomes is potentially large. If delayed specialization enables students to make better-informed decision about their field of specialization, the returns to education are affected through two channels. First, better-matched students are more likely to pursue careers in a field related to their studies, thus making better use of their specialized training. Second, workers who are in occupations which are well-suited to their innate talents are likely to be more productive. Education reforms which seek to increase the returns to college education must take account of such effects, particularly those reforms which would narrow the breadth of college education.

Does broad education help students discover their idiosyncratic talents? And if so, does the accuracy of this match translate into better labor market outcomes? This paper provides empirical answers to these questions. We first document the positive cross-sectional correlation between the timing of specialization and the probability of working in an occupation unrelated to college major. This finding suggests that selection into specialization needs to be accounted for, and motivates our structural model.

In our model, the optimal timing of specialization reflects a trade-off between identifying one's field of comparative advantage and acquiring specialized skills. Each agent is best suited to one field, but the identity of that field is initially unknown. Taking courses in many fields simultaneously provides agents with information about their comparative advantage. College course choices solve an optimal experimentation (bandit) problem: a student may choose multi-disciplinary education, where he acquires skills and receives information about his match to different fields; alternatively, he may choose specialized education, which conveys field-specific skills at a faster rate but does not provide such information.

Students update their beliefs about their comparative advantage by filtering a diffusion process in continuous time. Their course choices follow a stopping rule: students start by enrolling in multi-disciplinary education, where they learn about their field of comparative advantage. Once their confidence level is sufficiently high – the belief that they belong in either field being sufficiently close to 1 – they specialize. This bandit problem is not stationary: over time, as agents remaining in the mixed-education stream acquire skills, the value of their foregone wages rises and the expected length of additional specialized studies diminishes. This reduction in the length of specialized education depresses the value of current information and makes the agents less willing to experiment. Agents optimally lower the confidence level they require in order to specialize.

This property of optimal experimentation has important cross-sectional implications. Agents whose beliefs process drifts the fastest, and who therefore specialize early, are more likely to specialize in the field of their comparative advantage. Agents whose beliefs process remains close to their prior longer take their specialization decision on the basis of weaker information. Late specializers are therefore more likely to choose wrongly. Since these same individuals spent most of their studies in the multi-disciplinary stream, their education is also relatively more transferable. The model therefore delivers our reduced-form result: compared to early specializers, late specializers are more likely to change occupations.

To separate the contribution of self-selection from the lock-in effect of early specialization, the model is estimated structurally. Using data from a panel of college graduates, we estimate the parameters of the model through simulated method of moments. Detailed transcript data allow us to construct a proxy for the timing of specialization based on the course mix a student chooses in each period. We then simulate the model, selecting parameters to match the observed timing of specialization, occupation field choice, and wages.

Our estimated parameter values reveal that the benefits of flexible specialization are large. Time spent in mixed-discipline studies is informative, although imperfectly so. One year of exploratory college courses is as informative about comparative advantage as the entire pre-college period; nevertheless, despite considerable time spent acquiring information, only 53% of students major in the field of their comparative advantage. Furthermore, being type-matched to one's occupation is well-rewarded – workers employed in the field of their comparative advantage earn a 20% wage premium. While our estimates indicate a large, one-time switching cost equivalent to 1.5 years of wages, field-specific education is highly transferable: ten years after graduation, out-of-field education is remunerated at 90% the rate of field-related schooling. While the parameters are estimated simultaneously, variation in the timing of specialization and occupation choice appears to drive identification of the precision of signals, both prior to and during college. These values in turn, along with the earnings moments, pin down the parameters governing returns and switching costs.

These parameter values allow us to compare the current college system to a Europe-style counterfactual where specialization is imposed at college entry. While such a policy will be welfare-reducing by construction, imposing a timing of specialization may be a necessary practical or cost-saving measure on the part of an education system. Our estimates suggest that such a policy would have non-trivial consequences. We predict a modest change (a reduction from 24% to 20%) in the proportion of agents that pursue careers outside of their field of specialization, as the lock-in effect of early specialization counteracts the poorer information students have at the start of college. The change in the allocation of individuals across occupations is more substantial. In the early-specialization counterfactual, 70% of workers are employed in the field of their ex-ante comparative advantage, down from 77% in the benchmark. The average welfare cost of such a policy would be equivalent to reducing education stocks by half a year of specialized studies, or to reducing expected discounted lifetime earnings by 1.5%.

This paper contributes to the small literature on the timing of specialization in higher education by introducing a model of endogenous timing choice. Malamud (2010, 2011) assumes, just like we do, that broad education is informative about horizontal match characteristics. He compares labor market outcomes of early-specializing students (in England) to late-specializing students (in Scotland). Comparing these two cohorts, which face specialization imposed at different times, he finds that late specializers are less likely to choose occupations unrelated to their studies (2011 paper). Our model suggests that these results are reversed in a context where the timing of specialization is chosen by each student; a prediction which is borne out in our data. Flexible specialization times are typical of American colleges; our model therefore allows us to explore how the timing of specialization affects labor market outcomes in the US context.

Bordon and Fu (2013) estimate an equilibrium admissions model to explore the impact of unbundling college choice and major choice in Chile, where the current system requires students to apply to a college-major pair. The authors estimate the impact of alternate systems on college retention and peer quality, finding that more flexible policies are welfare improving so long as the relative returns to specialized education are not too high. Our model, which does not include college choice, allows us to estimate the returns to specialized education and the information value of unspecialized studies simultaneously.

This work also relates to the literature that integrates information revelation into models of college major choice. Altonji (1993) introduces a sequential model of college education where both aptitude and completion probabilities are ex-ante uncertain. Subsequent papers generally adhered to Altonji's approach of sorting majors by difficulty. Arcidiacono (2004) estimates a rich structural model with four majors, using grades as signals of ability and allowing college students to change majors, or drop out, between the early and later periods of college. While his results show that students do sort across majors based on ability, wage differencials between majors – large as they are – are not sufficient to explain this sorting. Taking advantage of the expectations data in their Berea Panel Study, Stinebrickner and Stinebrickner (2014), find that students switch out of majors in math and science as a result of learning that they perform less well in those subjects than they had anticipated. These findings are supported by Arcidiacono et al. (2012), Arcidiacono et al. (2013b) and Ost (2010).[1] We diverge from these papers by focusing on horizontal type discovery. We find that purely horizontal considerations matter a great deal: within our subset of majors – all of which are relatively high earning – a graduate working in the field of his comparative advantage earns significantly higher wages than an individual with the same education who is not well matched to his occupation.

Two recent papers extend the literature in this direction. Kinsler and Pavan (forthcoming) model ability as a two-dimensional vector – loosely corresponding to math and verbal skills – which is only fully revealed on the labor market. Silos and Smith (forthcoming) allow college students to choose their investment in three skills – quantitative, humanities and social sciences – which are ultimately rewarded to different degrees in different occupations. Residual uncertainty about agents' match to different occupations is resolved after a probationary period working in that particular job. In contrast to these papers, we are concerned with how education choices can help resolve this uncertainty prior to the labor market. This emphasis allows us to speak directly to education policy, unpacking the returns to higher education and informing the debate on college reform.

Finally, this paper speaks to the literature on the returns to education breadth. Dolton and Vignoles (2002) include secondary school course diversity in earnings equations. In their UK data, breadth of courses, at A-level or O-level, are shown to have insignificant effects on earnings. Using data from a European post-secondary graduate survey, Heijke and Meng (2006) find that graduates from programs that provide both academic and discipline-specific competencies produce less performant workers. There are many reasons to expect that selection into broad studies is not random, making a causal interpretation of these findings problematic.[2] Our paper contributes to this literature by exploring one specific channel through which broad curricula could be beneficial: by facilitating better field choices. In doing

[1]We do not treat college attrition in this paper; however, it remains an important focus of this literature. See Stinebrickner and Stinebrickner (forthcoming), Arcidiacono et al. (2013a) and Trachter (forthcoming), among others.

[2]To the best of our knowledge, no study has estimated the returns to breadth in education in a setting where education breadth is plausibly exogenous. Joensen and Nielsen (2009) come perhaps the closest, although the change in breadth they consider is quite marginal: taking advantage of a policy experiment in Denmark, which allowed students to take advanced math without taking advanced physics (taking advanced chemistry instead), the authors estimate the returns to advanced coursework in math.

so, we shed light on why broad curricula – even if they improve field choice – might be unrelated to labor market outcomes in the cross section.

In the following section, we introduce the data and describe how we determine the timing of specialization. Section 3 presents the model: first a two-type, continuous time version, and then the N-type, discrete time model we use in our simulations. The estimation procedure and the results are outlined in Section 4, while Section 5 discusses the policy experiments. Section 6 describes robustness exercises and extensions and Section 7 concludes.

## 2. Data

2.1. **Overview.** The data we use in this paper come from the restricted version of the Baccalaureate & Beyond 93:03 dataset.[3] This 10-year panel follows approximately 10,000 students who earned a bachelor's degree[4] from an American institution in the 1992-1993 academic year. Three follow-ups are carried out (one, four and ten years later), during which labor market and further education variables are collected.

Two features make this dataset ideal for our purposes: the policy context and the level of detail in the education and labor market variables. We require data in which there is some variation in the timing of specialization among students who graduate with the same major and degree. Given that college students in the US have considerable discretion over their course choices (and indeed, take relatively few courses in their major field), they have some scope to choose when they choose their major. This motivates our use of US data.

Our more restrictive requirement is that the panel include data on all courses taken in college, and the date at which these courses were taken. Timing of specialization, although central to our story, is not a well-established variable[5]

[3]Referred to in the following as B&B93:03. Dataset sponsored by the National Center for Education Statistics, U.S. Department of Education (for more information, see Wine et al. (2005)).

[4]A bachelor's degree is a 4-year undergraduate degree.

[5]The closest measurable variable might be the declaration of a major, as used in Bradley (2012)'s study of major choice during recessions. For a number of reasons this is not fully satisfactory for our purposes: first, different universities may require students to declare a major at different times, and this information is not in our data; second, if students are required to declare a major in order to, for example, register for their second year courses, this declaration is not necessarily an active form of specialization. Given that it is generally easy to change majors, and that major choice does not usually constrain course choices (at least early in college), the declaration of a

and we rely on the sequence of course choices to derive a proxy. Few datasets include both labor market outcomes and detailed course data.

We make several important sample restrictions. First, since we estimate information revelation in college, we retain only those students who graduated between the ages of 21 and 23:[6] older students are likely to have acquired information through other means, such as by working. Second, as our model supposes that students anticipate specific college courses to be useful on the labor market, we restrict our analysis to applied fields.[7] Finally, due to differences in pre-college ability across majors, as well the associated earnings differences on the labor market, we retain only the high-earning quantitative majors in our estimation.[8] Details of the sample construction can be found in Appendix A.

2.2. **Timing of specialization.** We derive a proxy for specialization using the mix of courses a student chooses in each period of school. In contrast with some panels that follow students through college,[9] our data does not include self-reported college major at multiple periods of time. Instead, we observe the entire sequence of course choices, as they appear on the final college transcript. Our specialization proxy in an indicator variable based on the share of credits chosen in the eventual major field of study. The intuition behind this is that students have a stronger incentive to take courses in the field in which they will eventually work. Once a student has settled on a particular field, the expected return from taking related courses rises.

Table 1 lists the total credits earned by students in each major, along with the share of those credits taken in that field. American bachelor degrees are evidently quite broad: even with coarse major categories, the share of credits taken in-major rarely exceeds 50%. We therefore do not expect students to commit themselves

major could be little more than a statement about the field a students thinks it is *most likely* she will pursue. Finally, this information is not available in the B&B93:03 dataset.

[6]These students in general took between 3 and 6 years to complete their degree.

[7]We are by no means the first to make a distinction between majors based on their links to the labor market. To cite two of many examples, Saniter and Siedler (2014) distinguish between fields with a strong versus a weak labor market orientation in German data, based on whether or not the field of study leads to a particular profession. Using Canadian data, Finnie (2002) finds differences in the early labor market outcomes of graduates from 'applied' fields versus those with majors in 'softer' subjects.

[8]These differences are explored in depth by Arcidiacono (2004); the distribution of SAT or ACT scores across majors for our sample is given in Figure 11 in Appendix A. We retain those majors with more mass in the upper two quartiles than in the lower two.

[9]For instance the NLS72, used by Arcidiacono (2004).

full-time to their major once they have specialized[10] – only that courses will be chosen differently after specialization than before.

TABLE 1. Credits: mean total and specialization by major

| **Major** | **Total credits** (SD) | **In-major share** |
|---|---|---|
| Science | 127 (13) | 0.51 |
| Engineering | 135 (12) | 0.43 |
| Business & Econ | 126 (11) | 0.43 |

Source: B&B93:03, sample restrictions described in section A.1. Note: a bachelor's degree requires 120 credits.

We derive the timing of specialization as follows. Courses are first associated with a major:[11] this allows us to calculate the total credits earned in each field for each academic year.[12] The timing of specialization is then defined as the term in which the share of credits taken in the eventual major field of study exceeds a given threshold.

To choose an appropriate threshold, we must take into account two factors: how strictly to define specialization, and how to treat differences across majors. If the threshold is too slack, it is difficult to justify that we are capturing a genuine change in course-choosing behavior; on the other hand, given the low number of credits students take in their major field, a high threshold results in many students 'never' specializing, despite the fact that they successfully graduate with a major.[13]

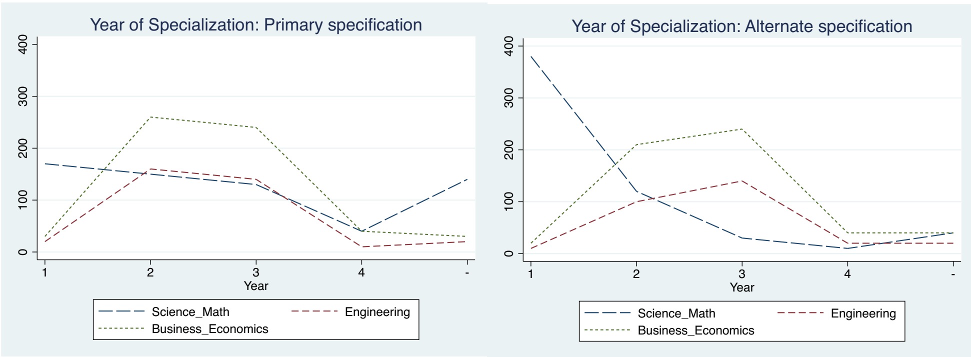
The major categories retained for analysis – while common and intuitive – are not necessarily aggregated in a similar fashion. Consider, for example, a major in science and a major in engineering. The field of science includes biology, physics, chemistry and math, and so courses in all these fields will be coded as science courses. In contrast, the field of engineering is relative narrow, since it includes

[10]There could be many reasons for this: it may not be possible to take a full load of courses in that field, due to missing course prerequisites or simply a shortage of courses at a given level; it may not be desirable to do so, particularly if advanced level-courses are more difficult than introductory courses in other fields; it may not even be permitted within the confines of the bachelor degree program, as many institutions impose a minimum number of courses to be taken in fields different from one's major.

[11]In the dataset, there are approximately 1000 unique course codes. We attribute each of these to one of 14 coarse major categories, several of which are later aggregated (see Table 19 in the Appendix).

[12]Courses vary in how long and intensive they are. The dataset includes a conversion metric for each course, translating the credit units attributed by the degree-granting institution into standard credit equivalents based a 120-credit degree.

[13]Graphical representations of sample retention and timing of specialization for increasingly strict thresholds are presented in Appendix A.3.




Major-specific threshold | 40% threshold

Figure 1: Timing of specialization: major-specific threshold vs. constant threshold

only course which are catalogued as 'engineering'. To address this issue, we adopt a data-driven approach to defining the threshold. Specifically, we choose a threshold for each major such that 50% of students who graduate in that field specialize by the end of their second year. Students who never reach the threshold, but who nevertheless graduate with a major in that field, are assumed to have specialized at the very end of college.

Figure 1 shows the distribution of the timing of specialization, for both our primary specification and for a universal 40% concentration threshold.[14] The two specifications give a similar picture for Engineering and Business & Economics majors, although the timing of specialization is slightly earlier using the major-specific threshold. Science & Math majors display two important differences. First, a 40% threshold is quite low for this group of majors: a vast majority of students meet this threshold in their first year.[15] Second, even using a major-specific threshold, students who complete a major in Science & Math tend to specialize earlier than students in other majors. This feature is robust to all of the timing of specialization specifications we have developed (see Appendix A.3 for a discussion of these approaches).

2.3. **Reduced-form evidence.** Table 2 presents a reduced form regression of the timing of specialization on the probability or working in an occupation related to

[14]This approach attributes specialization to the year in which the concentration of major-related courses meets or exceeds 40%.
[15]As discussed above, this is not surprising. A student taking 5 courses a semester must only take 2 of these in science, math or computer science to reach a 40% threshold.

one's field of study, ten years after college. While Malamud (2011) finds that early-specializing English students are more likely to change fields on the labor market than late-specializing Scottish students, our data display the opposite result: late specializers are more likely to work in a field unrelated to their studies. Restricting to our core sample of quantitative graduates, delaying specialization by one year is associated with a 1.5% decrease in the probability of working in an occupation related to one's major.

Our cross-sectional findings are echoed elsewhere. Using US data, Silos and Smith (forthcoming) also find that students with less specialized education stocks are more likely to switch occupations. Their concept of specialization, the closeness of a student's skill bundle to the average skill bundle of a given occupation, is different from ours; however, the concept of hedging through skill diversification is closely related. In a similar vein, Borghans and Golsteyn (2007) estimate a model of occupation changes where human capital is imperfectly transferable. Using Dutch data, in which almost 30% of graduates were working in an unrelated field 3 years after college, they find that higher skill transferability is associated with a greater probability that a graduate who regrets his field of study will switch to an occupation in a different field.

Table 2. Match probability

| | **Probability of working in related field** | | |
|---|---|---|---|
| | All majors | Quantitative | Non-quantitative |
| Timing | -0.0376*** | -0.0154** | -0.0244** |
| | (0.000) | (0.048) | (0.027) |
| Controls | X | X | X |
| $R^2$ | 0.381 | 0.527 | 0.771 |
| adj. $R^2$ | 0.375 | 0.522 | 0.765 |
| Sample size | 2110 | 1560 | 550 |

Source: B&B93:03, sample restrictions described in section A.1. P-values in parentheses; * $p < 0.10$, ** $p < 0.05$, ***. Sample sizes rounded to the nearest 10. Occupation is observed 10 years after graduation. Timing is the primary timing of specialization variable (see Section 2.2); match refers to the relation between field of study and field of work. Controls are major and occupation dummies.

Table 3 presents results of a regression of log income on the timing of specialization, controlling for field of study and occupation. The coefficient on timing of specialization in the main regression is small and not significant. While the contexts under consideration are quite different, this result is consistent with Dolton and Vignoles (2002)'s finding that curriculum breadth is unrelated to earnings.[16]

The absence of a trend in the cross sections hides interesting subgroup effects. When the sample is restricted to those individuals who switched fields on the labor market, timing becomes positive and significant: late specialization is associated with higher wages. If we consider only those individuals who are working in the field of their major, the coefficient on timing is negative but not significant.

Table 3. Log income

| | **Log income** | | |
|---|---|---|---|
| | All | Matched | Not matched |
| Timing | 0.00734 | -0.0131 | $0.0323^{*}$ |
| | (0.527) | (0.411) | (0.063) |
| Controls | X | X | X |
| $R^2$ | 0.189 | 0.191 | 0.219 |
| adj. $R^2$ | 0.182 | 0.185 | 0.203 |
| Sample size | 2100 | 1220 | 880 |

Source: B&B93:03, sample restrictions described in section A.1. P-values in parentheses; $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$. Sample sizes rounded to the nearest 10. Timing is the primary timing of specialization variable (see Section 2.2); match refers to the relation between field of study and field of work. Controls are major and occupation dummies.

Can these results be consistent with multi-disciplinary studies helping individuals discover their comparative advantage? A naive reading of Table 2 could conclude the opposite: after all, more time spent in broad education is associated with a higher probability of changing fields. If the timing of specialization is endogenous, however, the issue of selection looms large: those who are quite confident about their comparative advantage may opt out of multi-disciplinary studies at an early stage, especially if specialized education is more highly rewarded in their intended field of work. This selection pressure is intensified by the fact that the opportunity cost of studies is rises the longer a students spends in school. We are therefore unable to conclude on the learning value of multi-disciplinary education from cross-sectional regressions on the probability of switching fields.

The results in Tables 2 and 3 motivate us to develop a model where the timing of specialization is endogenous to the student's confidence about his comparative

[16]To see why, note that since late specializers will spend more of their degree taking a broad range of courses, the timing of specialization is correlated with college course breadth: late specializers chose broader curricula.

advantage. In Table 2 we observe a positive correlation between the timing of specialization and the probability of working in an occupation related to field of study. Table 3 shows that, among those workers who change fields, late specializers have higher earnings than early specializers. These results are consistent with the imperfect transferability of skills across fields: late specializers have a more portable skill set, and are therefore able to change fields more easily. They are also consistent with the gradual and imperfect revelation of information about students' comparative advantage, through a process of selection described in detail below. These two channels are indistinguishable in reduced-form evidence: to identify them separately, we will need to estimate the model structurally.

## 3. Model

3.1. **2-Type Continuous Time Model.** In this section we describe a two-field, continuous-time version of the model, which enables use to arrive quickly at the optimal policy and to introduce the parameters of interest. Restricting to two types helps convey the intuition and makes the learning process most transparent. The model that we estimate – with $N$-types and discrete periods – is described in Section 3.2.

3.1.1. *Agents.* There are two fields of work ($f \in \{s, a\}$), with two corresponding subjects taught at school, so human capital is a two-dimensional state variable $e = (e_S, e_A)$. Agents are born at date $t = 0$ with human capital $(e_S, e_A) = (0, 0)$, have an infinite lifetime and discount the future at rate $r > 0$. They have one of two possible comparative advantages (type $\theta \in \{S, A\}$).

3.1.2. *Choice set and education tenures.* At each point in time, agents can enter the labor force, follow specialized studies in subject S or in subject A, or follow multi-subject studies $M$.[17] The laws of motion of human capital are respectively $(\dot{e_S}, \dot{e_A}) \in \{(0,0), (1,0), (0,1), (1/2, 1/2)\}$. Since we assume that uncertainty is resolved upon entry on the labor market, we do not allow agents to return to school once they have started working.

[17]The idea that education has both general and specialized segments is echoed by Altonji et al. (2012), who explicitly model college as two decision periods: one where the student takes many courses, and a second where they choose a major.

3.1.3. *Information acquisition.* The type $\theta$ is initially unknown to the agent, who enters date 0 with a prior belief $p_0 = \mathbb{P}[\theta = S|\mathcal{F}_0]$. The prior $p_0$ may be correlated with $\theta$, and thus reflects information acquired prior to higher education.[18] Mixed-discipline education has an informational benefit: provided that he is engaged in multi-subject education for a short time interval $dt$, the agent observes an informative signal about his type, modeled as a diffusion $\tilde{Y}$ with type-dependent drift. [19] Filtering this observation allows him to update his estimate $p(t) = \mathbb{P}[\theta = S|\mathcal{F}_t]$ following Bayes' rule. The learning technology is characterized by a single signal-to-noise ratio $\phi$, such that the agent correctly forecasts his own belief as a pure drift-less diffusion (Beliefs derived from Bayes' rule are always martingales):

$$\mathrm{d}p(t) = p(t)\,(1 - p(t))\,\phi \mathrm{d}\tilde{W}(t) \tag{1}$$

The volatility of beliefs, roughly equivalent to to the speed of learning, is highest the closer the agent is to indecision ($p(t) = 1/2$) and the higher the signal-to-noise ratio. Learning is informative, so conditionally on type the belief process of $\theta = S$ types drifts upwards (similarly type $A$-agents' beliefs drift downwards):

$$dp(t) = (1 - p(t))^2 p(t)\phi^2 \mathrm{d}t + p(t)\,(1 - p(t))\,\phi \mathrm{d}\tilde{W}(t) \tag{2}$$

A more informative signal helps the belief converge towards the truth faster by increasing the drift of the process, but also raises its volatility.

This signal technology is characterized by gradual and incomplete learning: unlike in Poisson bandit models, agents never learn their type completely but instead continuously and gradually update their beliefs. An important property of this learning technology is that agents do not necessarily acquire better information over time.[20]

[18]$\mathcal{F}_t$ denotes time-$t$ filtrations which summarize the agent's information accumulated up to date $t$. In the binary case, a single scalar $\mathbb{P}[\theta = S]$ fully characterizes the beliefs of the agent at any point in time. With $N$ fields, a belief is identified with an element of the $N$-dimensional simplex.

[19]This standard model of gradual and continuous learning is used by Felli and Harris (1996); Moscarini and Smith (2001), among others.

[20]By contrast, filtering a sequence of Gaussian signals with unknown mean results in a Gaussian posterior belief about the mean, and the variance of the belief decreases deterministically. In that sense, agents necessarily become better informed over time, whereas in our setup, once an agent returns to a previously-held belief, it is as if any information received in the meantime had been wasted.

3.1.4. *Payoffs.* Payoffs accrue to agents from two sources: wages and benefits while working, and flow payoffs in school. During their education tenure, agents earn constant, possibly negative flow payoffs $z$ that reflect tuition and overall enjoyment of studies. At the end of the education tenure, agents choose a field of work which may differ from their field of specialization. We assume that the labor market rewards within-field education more than out-of-field education and to describe labor market returns, we first define an effective stock of skills for each field:

$$\epsilon_S = e_S + \beta e_A, \; \epsilon_A = e_A + \beta e_S. \tag{3}$$

While a unit of human capital in subject $s$ contributes one unit to $\epsilon_S$, an equivalent investment in subject $a$ contributes only $\beta$ to the stock of skills applicable to field $s$. We call $\beta$ a transferability parameter and assume $\beta \leq 1$: skills acquired in one field are only partially transferable. The flow wage when working in sector $f$ is given by:

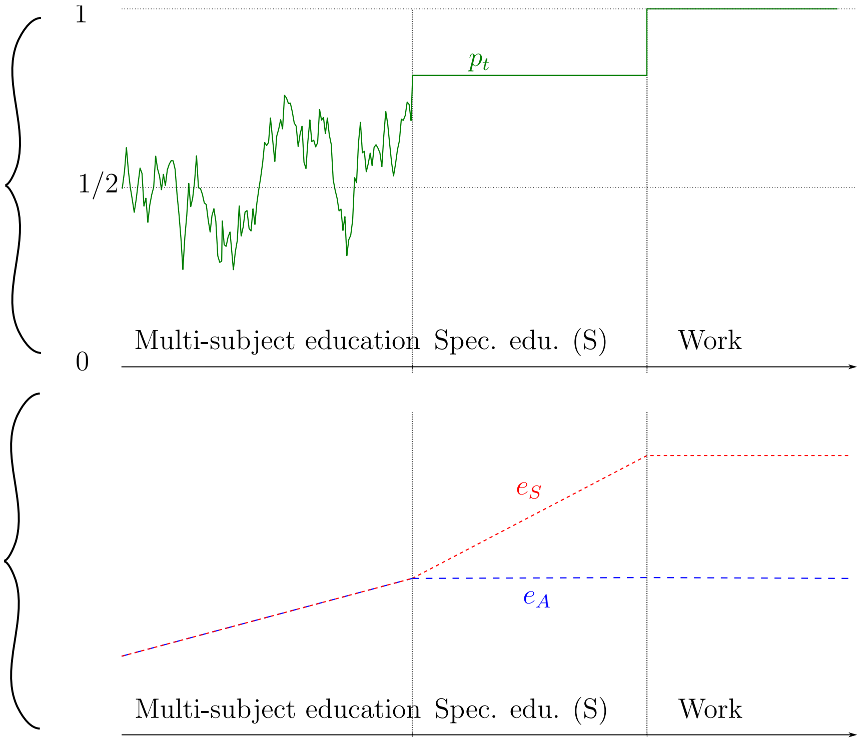
$$R(\epsilon_f) + \mathbb{1}_{\theta=f} P \;, \tag{4}$$

where $R(.)$ is an increasing and concave function and has the dimension of flow utility. Returns are therefore increasing in within-field skills ($R'(\epsilon_f)$ at the margin) and out-of-field skills ($\beta R'(\epsilon_f)$ at the margin), with an additive premium $P$ for working in the field of one's comparative advantage ($\theta = f$).

3.1.5. *Optimal policy and testable implications.* The optimal policy is characterized by optimal stopping times. Agents begin in multi-subject education, specialize their studies at time $\tau_1 \geq 0$, and proceed to the labor market at time $\tau_2 \geq \tau_1$.

It can be optimal for agents to choose $\tau_1 = 0$ if the signal is totally uninformative ($\phi = 0$) and more generally, experimentation may not be worthwhile. The length of the specialization period may also vanish ($\tau_2 = \tau_1$) if the return function does not depend on the level of effective human capital ($R'(\epsilon) = 0$), so that agents transfer directly from multi-subject education into the labor market. Figure 2 represents a regular case in which $\tau_1 > 0$ and $\tau_2 - \tau_1 > 0$.

The agent's decisions are presented recursively over the following paragraphs. Starting from the choice of occupation, we discuss the determination of the time


1
$p_t$
1/2
Multi-subject education
Spec. edu. (S)
Work
0
$e_S$
$e_A$
Multi-subject education
Spec. edu. (S)
Work

The agent's expected payoff at the end of his education tenure is $Y_s$, such that:

$$
\begin{aligned}
rY_s\left(p,e_S,e_A\right) = \\
& p\left(R_s\left(e_S+\beta e_A\right)+P\right) \\
& +(1-p)Max\left\{R_s\left(e_S+\beta e_A\right),R_a\left(e_A+\beta e_S\right)+P\right\}. \qquad (5)
\end{aligned}
$$

Indeed, with probability $p$ the agent is truly of type $S$, in which case he earns the premium $P$ when working in field $S$. With probability $1-p$, he is in fact type $A$: he may then choose either to remain employed in field $s$ and forgo the matching premium, or to switch fields. If he switches fields he does earn the premium; however, he suffers a transferability penalty since in his case $e_S+\beta e_A > \beta e_S + e_A$. Equation 5 allows us to define the optimal length of specialized studies, which solves the deterministic program:

$$
V_S\left(p,e_S,e_A\right) = \underset{\{H\geq 0\}}{Max}\int_0^H \exp\{-rt\}zdt + \exp\{-rH\}Y_s\left(p,e_S+H,e_A\right). \qquad (6)
$$

It follows that a student who specialized in subject $S$ will continue in that field until the marginal value of studies (flow value and increment in future earnings) falls below the opportunity cost of remaining at school (the flow-equivalent value of working). This is formalized in the following first-order condition:

$$
z+\frac{\partial Y_s\left(p,e_S+H,e_A\right)}{\partial e_S} \leq rY_s\left(p,e_S+H,e_A\right), \text{ with equality if } H>0. \qquad (7)
$$

Since this tradeoff is known at the time of specialization, the duration of specialized studies ($H=\tau_2-\tau_1$) is deterministic, as no uncertainty emerges during that time. Finally, since either field can be chosen as specialization, we can define a value of specializing $V=\max(V_A,V_S)$.

*Determination of* $\tau_1$. Due to uncertainty about the beliefs process $p(t)$, $\tau_1$ is not deterministic. While human capital accumulates steadily during multi-subject education (education stocks at time $t$ are $(t/2,t/2)$), beliefs evolve stochastically. The expected value writes as:

$$
\mathbb{E}\left[\int_0^{\tau_1}\exp\{-rt\}zdt+\exp\{-r\tau_1\}V\left(p(\tau_1),\tau_1/2,\tau_1/2\right)\right] \qquad (8)
$$

The expectation is taken over paths of the beliefs process $p(t)$ and $\tau_1$ is chosen optimally.

For each date $t$ we can define two boundary beliefs: one close to certainty in type-$S$ (call it $p_s(t)$), and one close to certainty in type-$A$ ($p_a(t)$). If, at time $t$, the agent's beliefs exceed the boundary in either direction, it is optimal for him to specialize. $\tau_1$ is then the first random time such that either $p(t) \geq p_s(t)$ or $p(t) \leq p_a(t)$. Figure 3 illustrates the boundaries, overlaid with a simulated belief path.

$$\tau_1 = Min\left\{t \geq 0, p(t) \geq p_s(t) \text{ or } p(t) \leq p_a(t)\right\} \tag{9}$$

Also pictured are the densities of exit times: the optimal specialization policy induces a distribution of exit times, with cumulative distribution $F_{ET}(t, \theta) = \mathbb{P}\left[\tau_1 \leq t | \theta\right]$. Similarly, by conditioning on the field of specialization and the agent's type, we can define a distribution of correct specialization times. For example, for $\theta = S$ agents, we can define $F_{CET}(t, \theta) = \mathbb{P}\left[\tau_1 \leq t \cap p(\tau_1) = p_s(\tau_1) | \theta = S\right]$. The limit of $F_{CET}(t, \theta)$ as $t$ grows large is the proportion of agents who specialize in the appropriate field.

3.1.6. *Properties of the optimal policy.* If we impose symmetry and assume that the returns to effective skills are linear, we can show that the specialization boundaries are bounded away from 0 and converge to 0 in finite time, and simulations indicate that optimal boundaries are monotonic. In Appendix B we illustrate the properties of simulated optimal policies and show how parameter changes affect the empirical predictions of the model; which shed light on the sources of identification in Section 4.

Monotonic boundary beliefs – with $p_s(t)$ decreasing and $p_a(t)$ increasing – have important empirical implications. They imply that early and late specializers have two important differences: not only do they accumulate different stocks of human capital, they also exit education with different probabilities of having specialized in the field of their comparative advantage. While all agents have the same information technology and use the same decision rules, differences in the idiosyncratic noise cause some agents' belief processes to drift rapidly towards the boundary, prompting them to specialize early relative to their peers. Consider two students

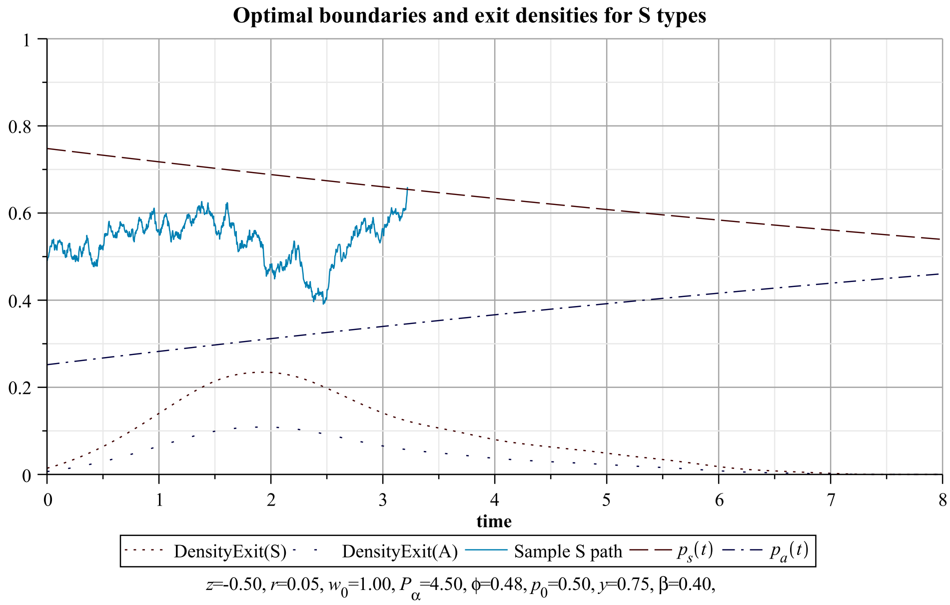



Figure 3: Optimal Boundaries and density of specialization times. *The agent starts specialized education once the belief process* $p(t)$ *escapes the interval* $(p_a(t), p_s(t))$. This particular sample path leads to specialization time $\tau_1 \approx 3.15$ and corresponds to correct specialization (in subject $S$).

that choose specialize in field $S$, the first at time $t$ and the second at time $T$ (with $t < T$). The first student specializes at a point where his belief satisfies $p(t) = p_s(t)$ while the second one's cutoff is reached as $p(T) = p_s(T)$. Since $p_s$ is a decreasing function, we have $p_s(t) > p_s(T)$: the first student specializes with a higher level of confidence than the second one. Furthermore, the belief level $p_s(t)$ is also statistically the cross-sectional proportion of types $S$ among agents specializing at time $t$. Fast learners thus are more likely to specialize *correctly* and be type-matched to their initial job field than slow learners.

Figure 3 illustrates one realized belief path of a single $S$-type student, overlaid with the optimal boundaries. Near $t = 0$, the agent requires his subjective probability of having comparative advantage $S$ to be above 0.71 or below 0.29 in order to specialize into fields $S$ or $A$, respectively. If he has not specialized by time $t = 7$, however, much smaller deviations from 1/2 would be sufficient to trigger specialization: beliefs above 0.55 or below 0.45 would suffice.

3.2. **Estimated N-type discrete time model.** To bring the model to data, we modify it in several ways that provide a better fit with the observations and respond to the computational challenges brought about by the rich informational structure. First, we estimate the college phase of the model in discrete time, which enables the use of standard dynamic programming techniques. Second, we consider a choice between three fields as opposed to two in the continuous time model, to accommodate the three high-ability applied majors that we retain for estimation. The modifications these changes entail are highlighted below.

3.2.1. *Computational challenges.* The curse of dimensionality affects bandit models particularly, as beliefs enter as multi-dimensional state variables in the optimization and the number of decision nodes increases very quickly, both with the number of periods and the with the dimensionality of beliefs (beliefs have $N-1$ degrees of freedom). This difficulty can be alleviated with the use of Gittins indices and recent papers[22] manage to accommodate binary state variables; however, non-stationary problems remain intrinsically challenging. Our formulation of the learning process postulates a sequential sampling of discrete signals (three per year of mixed education), which has the benefit of not necessitating further approximations. The discrete model limits the number of nodes at which the value function is estimated, first because the number of possible beliefs in the period following a given node is equal to the number of fields, but also because the ordering of signals is irrelevant, so several belief paths lead to the same belief. We retain the continuous-time formulation once agents specialize, which enables us to obtain explicit value functions in the specialization phase, and therefore simplifies and speeds up computation appreciably.[23]

3.2.2. *Agents.* There are $N$ fields of work (fields $f \in \{f_1, f_2, \ldots, f_N\}$), with $N$ corresponding subjects taught at school, so human capital is an $N-$ dimensional state variable $e = (e_1, e_2, ..., e_N)$. Agents are born at date $t = 0$ with human capital $(0, 0, \ldots, 0)$, have a finite lifetime $T$, and discount the future at a constant rate $\delta < 1$. They are endowed with an unknown comparative advantage in one of the $N$ fields (type $\theta \in \{1, 2, \ldots, N\}$). [24]

[22]See Papageorgiou (2014); Eeckhout and Weng (2011).
[23]Discount factors are adjusted to make preferences consistent.
[24]The model does not allow for heterogeneity among students other than their comparative advantage and the beliefs they hold about their comparative advantage. By limiting the number

3.2.3. *Choice set and education tenures.* Agents choose their education path as described in Section 3.1.2. For each period they are enrolled in multi-subject studies, agents now acquire $1/N$ units of education in each field.

3.2.4. *Information acquisition.* Agents begin higher education with an $N-$dimensional, type-dependent prior, $p_{0,\theta} = [P_1, P_2, \ldots, P_N]$. At the end of each period of multi-subject studies, the agent receive a noisy signal about his type.[25] Specifically, he observes a signal $\sigma \in 1, 2, ..., N$. With probability $\rho$, the signal corresponds to his type ($\sigma = \theta$), while with probability $1 - \rho$ the signal is misleading ($\sigma \neq \theta$) and correspond to any of the $N - 1$ other types. The agent updates his beliefs before choosing his education stream for the subsequent period. For example, if the agent holding belief $p$ observes $\sigma = 1$, his updated belief vector will be given by:

(10)
$$p' = \left[\frac{\rho P_1}{\rho P_1 + \gamma(1 - P_1)}, \frac{\gamma P_2}{\rho P_1 + \gamma(1 - P_1)}, ..., \frac{\gamma P_N}{\rho P_1 + \gamma(1 - P_1)}\right], \text{ with } \gamma = \frac{1-\rho}{N-1}.$$

3.2.5. *Payoffs.* Payoffs and occupational choice are as described in Section 3.1.4. In the $N$-type case, the effective stock of skills is defined as follows:

$$\forall n = 1 \ldots N, \; \epsilon_n = e_n + \beta \sum_{m \neq n} e_m. \tag{11}$$

There is no distinction across fields: all out-of-field education is treated symmetrically, as are all fields outside of an agent's comparative advantage.[26]

of majors considered, and by considering only successful college graduates, we narrow the span of ability within our sample; however, we acknowledge that vertical ability differences, not captured in our model, do remain. Table 23 in Appendix A.4 gives the correlation of the timing of specialization with several observable characteristics.

[25] A growing body of research explores how grades effect students' beliefs and course choices (see Arcidiacono (2004), Zafar (2011), Main and Ost (2014) and Stinebrickner and Stinebrickner (2014, forthcoming)). In this paper we remain agnostic about the source of the signals that students receive. While grades no doubt play a role, other unmeasurable factors also influence students' academic paths.

[26] We treat all 'out-of-field' education symmetrically. At the high school level, there is little evidence that any subjects are more universally rewarded than others (see Altonji (1995)). Math is something of an exception, though the evidence is sparse. Joensen and Nielsen (2009) find a strong causal effect of advanced math courses on later earnings in Denmark; elsewhere, the effect is small (Morin (2013), Canadian data), or only present in some groups (Levine and Zimmerman (1995), female college graduates in the US).

Flow returns (theoretical counterparts of log earnings) are given by:

$$y_f(\theta, (e_f, e_{-f})) = R(\epsilon_f) + \mathbb{1}_{\theta=f} P. \tag{12}$$

We briefly review the optimal behavior of agents and introduce the notation necessary to describe moment equations. We start by describing the last decision node, the decision to change fields, then we describe the optimal length of specialization and eventually the optimal experimentation phase.

3.2.6. *The decision to change fields.* We assume that type $\theta$ is perfectly revealed upon entry but agents face a cost $c$ of switching fields, a shortcut for the more realistic gradual realization of type mismatch and associated foregone experience and possible retraining.[27] This simplistic assumption also correspond to the limited observations of agents' early careers in the data. Agents who have stayed in mixed education until period $t$ and in specialized education in field 1 for $H$ years have education stock $(t/N + H, t/N, \ldots, t/N)$. Initially type-mismatched agents observe their type and the draw from the cost distribution $c$ and receive the following value if they switch to their field of comparative advantage $\theta \neq 1$:

$$J_{sw}(t + H, e, c) = -c + \int_0^{T-t-H} \exp\{-rt\} [R(\epsilon_\theta) + P] \, dt \tag{13}$$

If they remain in their field of specialization (field 1), they obtain value

$$J_{st}(t + H, e, c) = \int_0^{T-t-H} \exp\{-rt\} R(\epsilon_1) \, dt \tag{14}$$

They choose to switch fields provided that $J_{sw}(t + H, e, c) \geq J_{st}(t + H, e, c)$ or equivalently, whenever $c$ is lower than a cutoff value $\hat{c}(H)$.

From an ex ante perspective, upon starting work after training for a length $H$, initially type-mismatched agents obtain value

$$\bar{J}(t, e, H) = \int_{c_0}^{\hat{c}(H)} f(c) \, J_{sw}(t + H, e, c) \, dc + (1 - F(\hat{c}(H))) \, J_{st}(t + H, e, c), \tag{15}$$

[27] We do not model the labor market explicitly. This raises concerns that field-specific labor market fluctuations could affect our results, either by drawing in large numbers of students during booms or forcing graduates into other fields during crashes (in addition, Altonji et al. (2013) document that the returns to individual majors are affected differently by recessions). While we cannot rule this out, the sensitivity of our estimates should be reduced by the coarse aggregation of majors and occupations: large categories mean that those who cannot find work exactly corresponding to their major are likely to land a job in the broad field it is associated with. Furthermore, existing evidence suggests that the elasticity of major choice to market conditions, while positive, is relatively small: see Blom (2012) and Beffy et al. (2012).

where $f$ and $F$ are the *pdf* and *cdf* of the truncated exponential distribution. Agent who are type-matched to field 1 receive value

$$J_{tm}\left(t+H,e\right)=\int_0^{T-t-H}\exp\left\{-rt\right\}\left[R\left(\epsilon_1\right)+P\right]dt \tag{16}$$

3.2.7. *The optimal length of specialization.* At time 1, upon beginning specialization in field 1, the agent expects to be type-matched to field one with probability $p_1$ (the first entry in the beliefs vector) and type-unmatched with probability $1-p_1$. Given that the education stock and time are linked by the relation $e=(t/N,t/N,\ldots,t/N)$, we omit the explicit dependance on $e$. The expected value of specializing in field 1 writes as:

$$\begin{aligned}V_1(t,p)=\underset{\{H\geq 0,H\leq T\}}{Max}&\int_0^H\exp\left\{-rt\right\}zdt\\&+\exp\left\{-rH\right\}\left[(1-p_1)\bar{J}(t,e(H))+p_1J_{tm}(t+H,e(H))\right]\end{aligned} \tag{17}$$

Upon finding the maximizer $H^*(t,p)$, we can define the expected wage conditional on changing fields.

$$ew_{sw,1}(t,p)=R\left(\beta(t+H^*(t,p))+(1-\beta)t/N\right)+P \tag{18}$$

The expected wage conditional on staying in the given field is the weighted average of the wage of initially-matched and initially-unmatched agents. The total probability $A$ of carrying on in field 1 is the sum of the probability of being properly matched $A_1=p_1$ and the contribution of initially unmatched agents who remain in the field, $A_2=(1-p_1)\left(1-F\left(\hat{c}(H^*(t,p))\right)\right)$. Both groups receive as earnings $x_1=R\left(\beta(t+H^*(t,p))+(1-\beta)(t/N+H^*(t,p)\right)$ while type-matched agents receive the premium $P$.

$$ew_{st,1}(t,p)=(A_1(x_1+P)+A_2x_1)(A_1+A_2)^{-1} \tag{19}$$

3.2.8. *Optimal experimentation and beliefs histories.* Continuing in mixed-education enables agents to periodically receive signals about their field of comparative advantage, while increasing their stock of skills in all fields. In the discrete-time formulation, call $\delta$ the discount rate that can be compounded into the calibrated annual rate .[28] Recalling that a period is a third of a year, the Bellman equation

[28]There are three periods per year, so $\delta^3$ is the annual discount rate.

that defines the value of mixed-education reads:

$$V_0(t,p,e) = 1/3\, z + \delta\, \mathbb{E}\left[V(t+1/3, p', e')\right] \tag{20}$$

Where $V$ is the maximum of $V_0$ and the $V_i$ reflects optimal behavior from $t+1/3$ and $e' = e + 1/3 \times (1/N, \dots, 1/N)$. The expectation is taken over future values of $p'$, the bayesian update obtained as in equation (10).

The last object we define is the density of exit times, necessary for the definition of moments. Call $h_t(p)$ the $N$-dimensional vector such that the $i$-th entry reflects the mass of agents of type $i$ holding belief $p$ at time $t$, having never specialized before time $t$. Starting from a mass 1 of agents at date $-1$, we distribute them across types according to the vector $p_{-1}$ chosen so as to reflect the empirical distribution of majors, so $h_{-1}(p_{-1}) = p_{-1}$. Next, agents observe one signal which agrees with their type with probability $\rho_0$ and update their beliefs according to formula (10), replacing $\rho$ with $\rho_0$.

This procedure generates the time-0 beliefs $p_{0,i}$ that correspond to the Bayesian update of $p_{-1}$ upon observing signal $i$. To update the distribution, observe that type-1 agents receive signal 1 with probability $\rho_0$ while type-$i$, $i \neq 1$ agents receive signal 1 with probability $\gamma_0 = (N-1)^{-1}(1-\rho_0)$. The mass of agents of type 1 holding belief $p_{0,1}$ is therefore $h_0(p_{0,1})_1 = \rho_0 \times (p_{-1})_1$, where $(p_{-1})_1$ is the first entry of vector $p_{-1}$. The mass of agents of type $i > 1$ holding the same belief is $h_0(p_{0,1})_i = \gamma_0 \times (p_{-1})_i$, while the total mass of agents holding belief $p_{0,1}$ at time 0 is $\sum_{i=1}^N h_0(p_{0,1})_i$.

All possible belief points can be split between experimentation and specialization nodes. Say $p \in Ex(t)$ if agents holding belief $p$ at time $t$ choose to carry on in mixed-education and $p \in Sp_i(t)$ if they choose to specialize in field $i$ at time $t$. While experimentation prevails, we can update vector $h$ iteratively using the same procedure as above, except parameter $\rho$ is used in updates instead of parameter $\rho_0$.

## 4. Estimation

4.1. **Methodology.** We simulate the model described in Section 3.2. Parameter values are obtained through a combination of external calibration and simulated method of moments. The model is estimated using a subset of three majors: sciences & math, engineering, and business & economics.[29]

The parameters of the model are described in Table 4. While we have chosen to aggregate courses by year in the data, we estimate the model with shorter periods: each year is represented by three periods, with 12 periods being the maximum duration of non-specialized studies. This allows for a greater heterogeneity of beliefs and smoothness of decision nodes, reducing the granularity introduced through the discretization of the model.

For computational reasons, we impose that the cost of switching fields is drawn from an exponential distribution truncated[30] above a cutoff $c_0 > 0$ and we add the mean of the cost distribution ($C_\lambda$) to the list of estimated parameters. Since there is no mass on negative realizations of the cost, type-matched agents never find it advantageous to switch fields. Furthermore, the optimal stopping property and symmetry guarantee that type-unmatched agents do not find it optimal to switch to a field different from their comparative advantage.

4.2. **Moments.** The moments we use to estimate the model relate to four observables: the timing of specialization, the field of specialization, the relation between the field of work and the field of studies, and wages. Empirically, these correspond to four sets of moments. The first is the proportion of students specializing in each major, in each year (shown previously in Figure 1). These proportions fully describe the first two observables: that is, the timing of specialization and the field of specialization. The second is the probability of working in a field related to one's field of studies, conditional on major and timing of specialization. The third and fourth sets of moments are wages. We calculate two average wages for each major–timing of specialization cell: wages for those who are working in the field of their major, and wages of those who have switched to a different field.

[29] The inclusion of detailed major categories remains a challenge throughout the literature. In empirical work, Kinsler and Pavan (forthcoming) retain three majors (science, business and other), Arcidiacono (2004) uses four (natural sciences, business, education and social science/humanities/other), Stange (2013) includes three (business, engineering and nursing), while Altonji (1993)'s conceptual model has only two (math or science, and humanities). Our choice of majors is constrained by both computational power and cell size. While education majors are sufficiently numerous to be included, we restrict our primary sample to the more homogenous set of quantitative majors.

[30] The truncation enables to put probability mass on relatively high cost values without requiring a very low decay rate. The assumption of exponential distribution enables us to compute explicit continuation values without requiring dynamic programming in the specialized phase.

TABLE 4. Parameters

| **Parameter** | **Description** |
|---|---|
| Calibrated | |
| $p_{-1}$ | Prior belief: set to reflect the empirical distribution of majors* |
| $\delta$ | Annual discount factor: set to 0.96% |
| $T$ | Working lifetime: set to 20 years |
| $z$ | Flow cost of education: set to approximately -15% of average wage |
| Estimated | |
| $\rho$ | Precision of learning |
| $\rho_0$ | Precision of the initial signal |
| $\beta$ | Transferability |
| $P$ | Matching premium |
| $R_f(\epsilon_f)$ | Returns to education** with $R_f = R \; \forall \; f$*** |
| $C_\lambda$ | Expected cost of switching fields |

*This assumption is innocuous when the returns across fields are identical and there are no differences in flow utility from studying different subjects.

**Returns to effective education – the sum of in-field education and $\beta$ times out-of-field education – are assumed concave. We constrain this function to be a cubic polynomial with a value of one and a zero derivative at the maximum effective education level (the level of education corresponding to immediate specialization in any field for 4 years), and to be equal to zero at zero education. We estimate one remaining curvature parameter from the data.

***Returns do vary across fields, both in the cross-section (Carnevale et al. (2012) present recent evidence from the United States, Finnie (2002) from Canada, to name just a few) and in studies controlling for selection (e.g. Chevalier (2012), Kinsler and Pavan (forthcoming)).

The vector of moments therefore has 60 entries: 4 moment types, 5 years[31] and 3 fields. To formalize notation, let $MT$ be the set of theoretical moments, each specific to a year-field cell. The first block of moments concerns timing: entry $y + 3 \times (f - 1)$ is the proportion of agents specializing in field $f$ and year $y$:

$$MT_{y+3\times(f-1)} = \sum_{t=3(y-1)}^{3y-1} \sum_{p \in Sp_f(t)} \sum_{i=1}^{N} (h_t(p))_i. \tag{21}$$

The innermost summation symbol denotes the fact that we are summing across agent types, the intermediate one corresponds to the pooling of belief nodes leading to specialization in field $f$ in period $t$, and the outermost summation aggregates 3 periods, to bring the unit of observation from a period up to a year. Since all agents eventually specialize, the sum of the first 15 entries equals the total population.

[31]Recall from Section 2.2 that, although all students in our sample graduated with a major in one of the three fields, some never reach the threshold of specialization. We interpret this as *very late* specialization; that is, specialization just prior to entering the labor market.

The second block of moments measures the number of horizontally matched agents: entry $15 + y + 3 \times (f - 1)$ is the proportion of agents who specialize in field $f$ and year $y$, eventually carry on working in field $f$. Each such entry is the product of the corresponding timing entry and the year-field cell's average probability of remaining in the chosen field:

$$MT_{15+y+3\times(f-1)} = \sum_{t=3(y-1)}^{3y-1} \sum_{p \in Sp_f(t)} \sum_{i=1}^{N}$$

$$(h_t(p))_i \times [p_f + (1 - p_f)(1 - F(\hat{c}(H^*(t, p))))] . \tag{22}$$

These first two blocks of moments are proportions which are directly comparable to their empirical counterparts. The next two blocks represent earnings: entry $30 + y + 3 \times (f - 1)$ is the average earnings of agents who specialize in field $f$ and year $y$ and eventually carry on working in field $f$, while entry $45 + y + 3 \times (f - 1)$ is the average earnings of agents who specialize in field $f$ and year $y$ but work in a field different from $f$. These vectors are computed according to section 3.2.7 and averaged using the relevant proportions.

4.3. **Distance criterion.** In order to relate the earnings moments to their theoretical counterparts, particularly regarding orders of magnitude and dispersion, we first standardize them: standard deviations of income are therefore compared to standard deviations of returns in the model.

Our parameter estimates are the values which minimize the weighted difference between the empirical and theoretical versions of the moments described above. The model is simulated for every element $s$ in the set $S$ of parameters on a defined grid, and the estimates are the parameter values $\hat{\Theta}$ that satisfy:

$$\hat{\Theta} = argmin_{\Theta_s}(M_E - M_T(\Theta_s))'W(M_E - M_T(\Theta_s)), i = s...S; \tag{23}$$

where $M_E$ are the empirical moments, $M_T(\Theta_s)$ their simulated counterparts using parameter values $\Theta_s$, and $W$ is a weighting matrix. There are two issues with the determination of matrix $W$. The first one has to do with commensurability of measurements. Empirical and theoretical income measurement units must be made comparable, and then adjusted so that they are of similar magnitude to proportion-based moments. This is required so that no one set of moments dwarfs another, or the distance function would emphasize them disproportionately. To address both difficulties, we standardize sets of moments block-wise. That is, each block of 15 entries is adjusted linearly so as to have mean 0 and standard deviation 1.[32] Finally, we constrain our estimates to generate a pattern of majors consistent with the observed distribution – and thereby consistent with the prior. We do this by eliminating parameter values that cause the predicted share in each major to deviate from the observed share by more than 25%.

The weighting matrix used for the primary estimation is a diagonal matrix of the empirical proportions of the sample in each major-timing of specialization cell, that is, the first 15 entries of the $M_E$ vector, repeated 4 times. We also compute distance criteria that put a lower weight on income moments, to reflect the higher uncertainties associated with the modelling of earnings. Distance functions using the Identity matrix for weights, as well as alternative measurements of the timing variable and selected subsamples, are explored in Section 6.1.2.

4.4. **Estimates.** The values of the six estimated parameters are given in Table 5. These results support the existence of learning in general education and also imply that specialized education is imperfectly transferable across fields. Specifically, we find that each year in multi-subject information gives a signal with precision-level 0.39. This value can be compared to a pure noise 'signal', which in our 3-type case would have a precision of 0.33. Conversely, if a single period of broad studies revealed type with certainty, the precision of the signal would be 1. Our estimated signal is therefore informative, but still noisy.

Another way to understand the estimated precision level is by considering the *expected entropy reduction*, a pure measure of informativeness.[33] We find that the initial distribution of majors is associated with an entropy of 1.58, which the

[32]Populations (cell sizes) are used as weights in the standardization.

[33]If variable $\theta$ can take any of $N$ values, write $q_k = \mathbb{P}[\theta = k]$. Shannon entropy measures the uncertainty associated with belief vector $q$ and is defined as

$$E(q) = -\sum_{j=1}^{N} q_j log_2(q_j), \text{ with } 0 \times log_2(0) = 0 \text{ by convention} \tag{24}$$

To define expected entropy reduction, suppose the agent starts a given period with belief $p \in \Delta_N$. The signal $\sigma$ has $n$ possible realizations and with probability $\omega_i$, the bayesian update of $p$ following observation $\sigma_i$ takes value $g_i$. We can thus define the expected entropy of the posterior, leading to the following definition

$$I(\sigma, p) = E(q) - \sum_{i=1}^{n} \omega_i E(g_i) \geq 0 \tag{25}$$

pre-college signal reduces by 0.075. Each period of multi-disciplinary college (a third of a year) reduces entropy by a further 0.027. Students therefore acquire as much information in one year of broad college courses as they did in the entire pre-college period $(3 * 0.027)$.[34]

The parameter $\beta$ is estimated at 0.90; this implies that out-of-field education is remunerated at 90% of the level of education related to one's field of work. An individual suffers a modest loss of human capital when choosing to work in a field different from his major – more substantial if he specialized early. This loss of human capital is compensated for by a large premium to working in the field of one's comparative advantage. Our estimated matching premium is 0.20: those who are type-matched to their field of work earn 20% higher wages than similarly-skilled individuals who do not. Finally, students incur large one-time costs when they switch fields, equivalent to 1.59 years of income.

Table 5. Estimated parameter values

| **Parameter** | **Definition** | **Estimate** | **Discussion** |
|---|---|---|---|
| $\rho$ | Precision of learning | 0.39 | Compare to an uninformative signal: $\rho = 0.33$ |
| $\rho_0$ | High school signal | 0.49 | The precision of beliefs at college entry |
| $\beta$ | Transferability | 0.90 | Out-of-field education is remunerated at 90% of in-field education |
| $P$ | Matching premium | 20 | 20% percent of earnings are due to type-match with occupation |
| $R_f(\epsilon_f)$ | Return function | 23.5 | Curvature parameter for the returns function (no intuitive interpretation) |
| $C_\lambda$ | Expected switching penalty | 1.59 | Corresponds to ∼1.5 years of income |

4.5. **Model fit.** We explore model fit in two ways. First, we present graphically the relative and absolute deviations of the 60 theoretical moments from their empirical counterparts. The left panel of Figure 4 shows the contribution each

In investment problems, expected entropy reduction is shown by Cabrales et al. (2013) to be the unique parameter-independent complete ordering of information structures that agrees with investors' willingness to pay. It is therefore a valid measure of the informativeness of a signal, particularly when it comes to comparisons.

[34]Neither signal reduces the absolute value of entropy by a large amount, but entropy is a concave function of beliefs and decreases fast near the edges of the simplex, so absolute variations near the middle of the simplex are small.

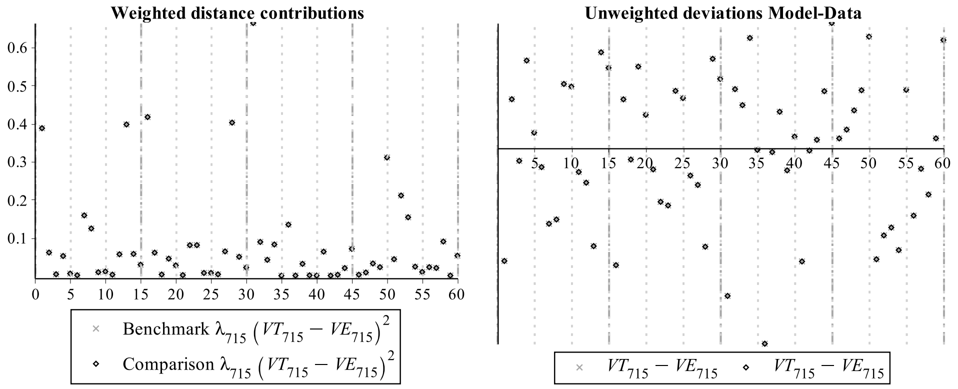
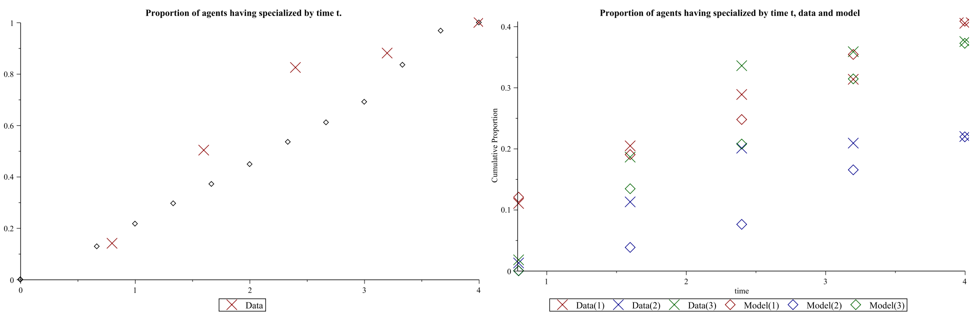



Figure 4: Relative (L) and absolute (R) deviations: model vs data

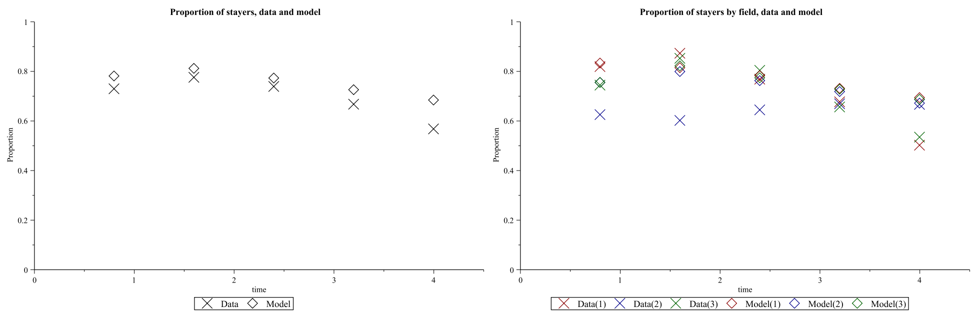



Aggregated By major

Figure 5: Model vs. data: specialization times

moment makes to the distance function (moments are numbered from 1-60, as described in Section 4.2). The absolute deviations behind these contributions are shown in the right panel of Figure 4.

To understand the implications of these differences, we next examine each block of moments individually. Figure 5 shows predicted and observed patterns of specialization. In the left panel, all majors are aggregated: black diamonds represent predicted specialization at different belief nodes, while red crosses plot the empirical counterpart. The right panel displays the same data, broken down by major (majors are (1) Science & Math, (2) Engineering, (3) Business & Economics). The model predicts that students will specialize slightly later, on average, that they do in the data. When broken down by major, we can see this arises primarily from a failure to match the mass of specialization by engineering students in periods 2 and 3, although the model under-predicts mid-term specialization overall as well.



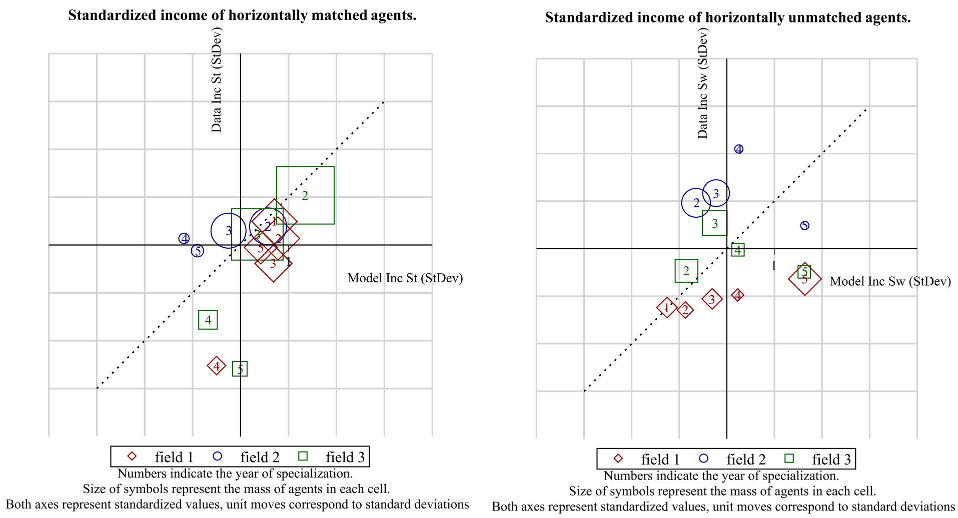

Aggregated By major

Figure 6: Model vs. data: probability of switching

The probability of working in the field of studies, given the timing of specialization, is shown in Figure 6. On aggregate (left panel), staying is over-predicted by the model. The disaggregated comparison (right panel) suggests that the low propensity of engineers to remain in engineering is driving the divergence between the model and the data. This is compounded for late specializers by a low probability of staying for the other majors as well.

The two sets of wage moments are presented in Figure 7, with stayers in the left panel, and switchers on the right. The graphs present standard deviations from mean income for both the model, on the horizontal axis, and the data, plotted on the vertical axis. If the model perfectly predicted income differences, the plotted observations should be arranged along the 45-degree line. Figure 7 shows that the model predicts income quite well for workers who remain in the field of their major. While there are some off-diagonal observations, these are generally small masses of individuals.

This is not so much the case for the income of switchers: in the data, engineers who switch fields earn systematically higher incomes, while science and math majors who switch fields earn low incomes. The model, which predicts lower incomes for early specializers who switch, matches empirical wages for early-specializing scientists and business & economics majors, but performs poorly elsewhere. This may be partly due to our coarse treatment of occupations outside of the field of study: in the model, all occupations unrelated to the field of study are treated symmetrically. Table 18 in Appendix A.2 lists the occupations of graduates from each major. A full 43% of all engineering graduates who switch fields are employed in business and management (a category which includes high-paying occupations




Stayers Switchers

Figure 7: Model vs. data: wages of stayers (L) and switchers (R)

in finance), whereas science and math graduates who are working in other fields are spread out across a greater diversity of occupations, including 29% employed in the low-paying field of education.

4.6. **Identification.** Although all parameters are estimated simultaneously, we can give some intuition about the identification process. We do so using two approaches. First, we compute illustrative simulated comparative statics for a linear-symmetric case of the model. These results are reported and discussed in Appendix C. Second, we constrain the parameters of our simulation one at a time, and re-estimate the remaining parameters. To accomplish this, we take the grid of parameter values that we used to estimate the model and impose the value of one parameter at a time. We then select the set of parameter values which, while respecting our imposed constraint, minimizes the primary distance criterion. In order to compare high and low values of the parameter in question, we do this for the highest and lowest value that parameter takes on our grid.These experiments suggest that the timing of specialization and switching behavior are primarily responsible for identifying the precision of learning, both prior to and during college. These parameters, along with the earnings moments, in turn pin down the parameters governing returns and switching costs.

The resulting parameter estimates are given in Table 6, and will be discussed in more detail in Section 6.1.1. We focus here on how these experimental variations affect the distance criterion, and the implications this has for identification. Table 7 gives the distance criterion for each experiment, broken down into the contribution of each block of moments. These block fall naturally into two groups: the first two capture the behavior of agents (major & timing of specialization, and the share of each major-timing cell who stayed in their field of study on the labor market), while the second two are wage moments (for stayers, and for switchers). The contribution of each of the 60 individual moments to the distance criterion is presented graphically in Appendix D.

4.6.1. *Learning parameters.* The first two rows of Table 7 impose the precision of signals received during multi-disciplinary studies. Imposing an imprecise signal actually improves the match with the empirical wage moments; however, it does so by worsening the match with the behavioral moments considerably. Imposing a high precision of learning has a smaller effect on the distance criterion, with the largest deviation from the benchmark coming from the wages of stayers. Overall, it appears that variations in the precision of the college signals during affects the distance criterion primarily through the behavior moments.

Imposing a highly informative pre-college signal impacts the distance criterion in a similar way as did the imposition of an *un-informative* college signal, and vice-versa for a highly informative pre-college signal: notice the symmetry between the first and second pair of rows in Table 7. This suggests that the precision of the pre-college signal is also being pinned down by the behavioral moments, although in this case the variation in the wages of stayers is also quite important.

4.6.2. *Return function parameters.* The next four rows of Table 7 present variations in the transferability of specialized education and the matching premium. Based on the distance criteria alone, it appears that these two parameters are relatively unimportant. In the case of the comparative advantage premium, $P$, the distance criteria for the low and high values are almost identical, and the contribution of each block of moments barely changes. In both cases, the distance criterion is relatively evenly contributed to from each block of moments.

A look ahead to the distance-minimize parameter values estimated under each constraint, listed in Table 6, suggests that these parameters matter a great deal. This is particularly the case for the estimation of the switching cost and the transferability of specialized education, which appear to move together: low switching costs coexisting with high transferability, for an 'easy mobility' alternative, and vice-versa for a 'tough mobility' alternative. The fact that these two very different alternatives have such similar distance criteria suggests that the identification of these parameters is not as strong as the others.[35]

4.6.3. *Switching costs.* The switching cost parameter appears to be driven by both the behavioral moments and the wage moments (see the final rows of Table 7). While the differences in each case are modest, with the experiment imposing high switching costs matching three of the four moments better than that with low switching costs, variation in this parameter appears to affect all four sets of moments to a similar degree.

Table 6. Estimated parameters when constraining one parameter at a time (actual estimates obtain by dividing by 100)

| Contraint | $\rho$ | $\rho_0$ | $P$ | $R_f(\epsilon_f)$ | $\beta$ | $C_\lambda$ |
|---|---|---|---|---|---|---|
| **Benchmark** | **39** | **49** | **20** | **23.5** | **90** | **1.59** |
| Low $\rho$ | **37** | 45 | 24.5 | 23.5 | 90 | 2.97 |
| High $\rho$ | **43** | 47 | 17 | 23.5 | 66 | 5.99 |
| Low $\rho_0$ | 40 | **39** | 15.5 | 23 | 70 | 5.68 |
| High $\rho_0$ | 39 | **51** | 24.5 | 23 | 90 | 1.2 |
| Low $P$ | 40 | 45 | **15.5** | 23.5 | 82 | 4.72 |
| High $P$ | 39 | 49 | **24.5** | 23 | 90 | 2.94 |
| Low $\beta$ | 41 | 49 | 18.5 | 23.5 | **66** | 6.82 |
| High $\beta$ | 39 | 49 | 20 | 23.5 | **90** | 1.59 |
| Low $C_\lambda$ | 39 | 47 | 15.5 | 22.5 | 90 | **1.11** |
| High $C_\lambda$ | 40 | 43 | 24.5 | 23.5 | 66 | **8.75** |

## 5. Policy simulation: the costs of imposing early specialization

Calls to reform college education in the US regularly accuse bachelor degrees of being too broad and weakly linked to the labor market. What would happen if students were forced to specialize at college entry? Using the parameter values estimated above, we can predict the impact of such a policy. Specifically, we

[35] We do not present experimental variation in the curvature of the returns function. While this parameter is important, it is not a primary focus of our study: we estimate it because we lack any reasonable outside calibration.

Table 7. Contribution of each block of moments to the distance criterion, under different constraints

| Constraint | Block 1 | Block 2 | Block 3 | Block 4 | TOTAL |
|---|---|---|---|---|---|
| **Benchmark** | **0.67** | **0.53** | **0.54** | **0.97** | **2.72** |
| Low $\rho$ | 1.43 | 1.41 | 0.34 | 0.96 | 4.15 |
| High $\rho$ | 0.70 | 0.62 | 0.84 | 0.86 | 3.01 |
| Low $\rho_0$ | 0.75 | 0.60 | 1.09 | 0.88 | 3.32 |
| High $\rho_0$ | 1.49 | 1.38 | 0.50 | 0.97 | 4.34 |
| Low $P$ | 0.75 | 0.61 | 0.49 | 0.90 | 2.76 |
| High $P$ | 0.72 | 0.61 | 0.46 | 0.97 | 2.77 |
| Low $\beta$ | 0.71 | 0.71 | 0.63 | 0.86 | 2.91 |
| High $\beta$ | 0.67 | 0.53 | 0.54 | 0.97 | 2.72 |
| Low $C_\lambda$ | 0.97 | 0.82 | 0.64 | 0.96 | 3.39 |
| High $C_\lambda$ | 0.86 | 0.75 | 0.72 | 0.83 | 3.15 |

Moment blocks: (1) timing of specialization and major, (2) stayers, by timing and major, (3) wages of stayers, (4) wages of switchers.

consider a policy where students must specialize after receiving a single college signal: this corresponds to spending 1/3 of a year in mixed-discipline studies prior to specializing.

We focus on two outcomes, summarized in Table 8. The first, which is observable in the data, is the fraction of students who choose an occupation in a field different from their major.[36] In our baseline simulation we find that 47% of workers are type-mis-matched to their field when they reach the labor market; nearly half of these workers, 24% overall, switch occupations and thus end up working in a field unrelated to their field of studies.[37] In the counterfactual, students do not learn about their type in college, and 50% graduate in a field different than their comparative advantage. These students specialized early, acquiring a large stock of human capital in their major field. The high transferability and large comparative advantage premium nevertheless induce many students to seek out their preferred field: 20% change fields when their type is revealed.

The second outcome is the number of agents who are working in a field different from that of their comparative advantage. In the baseline simulation, 53% of

[36]The large number of young people working in fields unrelated to their field of study has been studied in a number of countries. See, for instance, Finnie (2001) in Canada, McGuinness and Sloane (2011) in the UK, Badillo-Amador et al. (2005) in Spain, Bender and Heywood (2011) for scientists in the US.

[37]Although our criteria for matching occupations and majors is based on coarse categories, the level of horizontal mis-match in our data is similar to that found through other methods. Using the 1993 Survey of College Graduates, Robst (2007) finds that 20% of respondents – across all ages and majors – report that their work is 'not related' to their degree field.

students correctly discover their type during college: all of these are correctly type-matched on the labor market. In addition, 24% switch fields, leaving 23% of the population mis-matched to their field of work. Under the counterfactual policy, only 50% of students graduate in the field best suited to them; 30% overall remain type-mismatched to their occupation.

Table 8. Policy Experiment

| | Initially matched | Change fields | Remain type-mismatched |
|---|---|---|---|
| Baseline | 53% | 24% | 23% |
| Counterfactual | 50% | 20% | 30% |

This counterfactual experiment highlights the deep implications of our results for education policy. Imposing early field choice actually improves the correspondence between field of study and field of work: 17% fewer students choose an occupation outside their field of study. This apparent improvement masks a significant worsening of the allocation of individuals to occupations that suit them best: early specialization increases type-occupation mis-match by 30%. Our estimates suggests that the average individual cost of this policy is equivalent to the return earned on 0.45 of a year of occupation-related specialized studies, or approximate 3% of wages.

How does a student's expected value evolve as function of imposed timing of specialization? Figure 8 shows the ex-ante expected value for a range of mandated specialization policies. Note that the policy under consideration is specialization imposed *at or before* the date on the x-axis; prior to mandated specialization, students may opt in to specialization at any time. The left-most observation corresponds to the policy described above. The relationship between the mandated specialization time and the expected value is almost linear: while any constraint makes students on average worse off, the time at which specialization is imposed has a large impact.

We do not draw conclusions on whether or not early specialization is an efficient policy choice. There are two reasons for this. First, we do not have data on the relative costs of broad and specialized education. Anecdotal evidence suggests that the breadth of courses and flexibility of course choices at American universities presents non-trivial administrative challenges: early-specialization is often associated with simpler, homogenous course schedules. Second, we make the

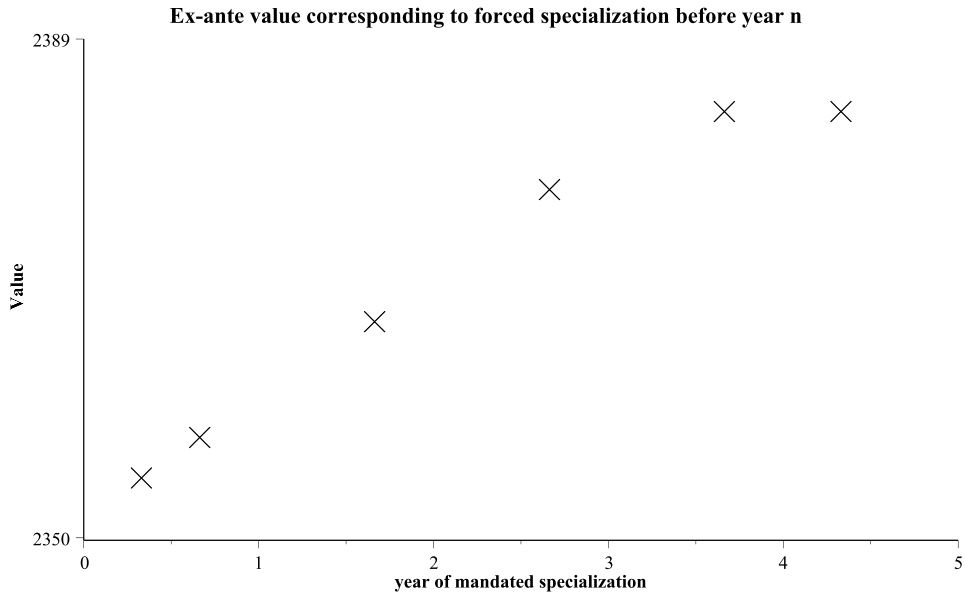



Figure 8: Ex-ante expected value under different specialization regimes

standard assumption that students know more about themselves than the social planner, and that they both process information optimally and make optimal experimentation decisions. It follows from this assumption that *any* constraint on course choices is at least weakly welfare decreasing. While our counterfactual will therefore make students worse off by construction, the magnitude of the effects we find can inform policies which take both the costs and benefits of allowing flexible course choices into account.

## 6. Robustness Checks and Extensions

6.1. **Robustness checks.** To explore the sensitivity of our estimates to individual characteristic which are outside our model, we perform three types of experiments. First, we constrain our parameter values one at a time, and estimate the remaining parameters. Second, we estimate the model parameters using different distance functions. Third, we estimate the model using different subsamples of the data.[38]

[38]For each subsample, the timing of specialization variable is attributed as it is in the primary estimation; however, the moments are adjusted to reflect the different population under consideration. The calibrated parameters are held constant across the subsample estimations, while the estimated parameters are allowed to vary. This means that the prior belief, which is calibrated to the empirical distribution of graduates in the full sample, is maintained for each subsample estimation. For this reason, we do not report estimates for subsamples with very different patterns of specialization from the full sample: doing so would violate the fixed-point assumption behind our prior beliefs.

While the interpretation of these parameter estimates is quite limited, they shed some light on the sensitivity of our results to different specifications. Overall, the picture is encouraging. The informativeness of mixed-discipline education varies little across the estimations; however, transferability of education and the matching premium are more volatile.

6.1.1. *Constraining parameters.* In Section 4.6 we introduced a series of experiments where we constrain the value of one parameter and estimate the remaining five. In addition to shedding light on identification, these experiments allow us to explore the robustness of our parameter estimates.

The first four experiments, reported in Table 6, concern the precision of learning before and during college. We first constrain the precision of learning during multidisciplinary studies to be low. The resulting parameter estimates, with respect to our baseline specification, have a lower level of pre-college information, but higher switching costs and a higher comparative advantage premium. If learning is imprecise, there must be high returns to making a correct match in order to justify observed behavior - and even then, this set of parameters matching the behavioral moments quite poorly (see Table 7. When imposing a high precision of learning, on the other hand, we find low transferability, high switching costs and a low premium. If learning happens quickly, students must expect that readjustment on the labor market is very difficult, otherwise they would not spend so much time acquiring information.

Constraining the precision of pre-college learning has the reverse effect: imposing a precise pre-college signal leads to estimates with high transferability and low switching costs, while imposing a highly noisy college signal leads to estimates with a high switching costs and low transferability. That the precision of college signals is also different in the two sets of parameters, with a precise pre-college signal associated with a less-precise college signal and vice versa, could partially explain this result. Notice, however, that precise pre-college signal condition – with a loose labor market – misses the behavioral moments quite badly.

Comparing the low and high transferability experiments[39] echoes the comparisons above by suggesting the existence of easy-mobility and tough-mobility alternatives. Imposing low transferability leads to a set of parameters with a *high*

[39]High transferability also corresponds to the benchmark.

switching cost, while high transferability co-exists with relatively low switching costs. Note also that the college signal is quite precise under the low transferability constraint. Variations in the comparative advantage premium perform similarly to variations in transferability with, as discussed in Section 4.6, a negligible difference in the distance function between the high premium and low premium conditions.

The final two rows of Table 6 compare parameter values when we impose a high cost of switching, and when we impose a low cost. As we have seen previously, to justify a low switching cost the transferability needs to be high, while the opposite is true for a high switching cost. In keeping with previous findings, the easy-mobility alternative is associated with less precise college signals and a stronger pre-college signal, although modestly in both cases.

These experiments suggest the existence of an alternate set of estimates, which may not be too distant from our best-fit parameters, with lower transferability, high average switching costs, and more precise college signals than our current benchmark.

6.1.2. *Alternate distance criteria.* In addition to our primary distance criterion, we consider several other distance functions. We consider six variants in two families of distance functions: the first, in keeping with our primary specification, weights each moment by the fraction of the sample found in the corresponding major-timing of specialization cell. This approach puts more weight on cells that are heavily populated. The second family of distance functions is not weighted,[40] but is otherwise identical to the first.

Table 9 presents distance-minimizing parameter estimates for the weighed distance functions, while Table 10 displays the unweighted equivalents. The primary estimates are listed in the first column of Table 9, for comparison. The distance functions differ in the moments which are targeted, and whether or not the moments have been standardized. These moments are, by column: (I) major & timing of specialization, probability of switching and wages, all standardized; (II) major & timing of specialization, probability of staying, standardized; (III) major & timing of specialization, probability of staying, not standardized; (IV) cumulative density of major & timing of specialization (as opposed to cell shares),

[40] The unweighted distance function uses the Identity matrix as the weighting matrix.

probability of match as proportion of cell (as opposed to share of population); (V) cumulative density of major & timing of specialization; (VI) probability of match as proportion of cell.

TABLE 9. Distance-minimizing parameters using alternate criteria - weighted

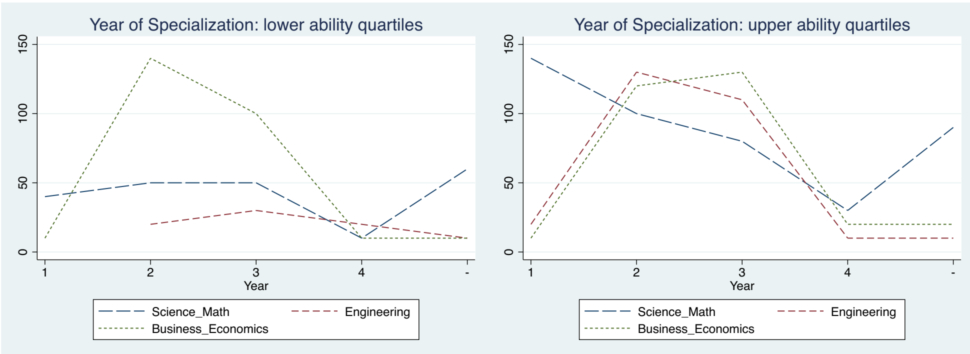
| | | Distance criteria | | | | | |
|---|---|---|---|---|---|---|---|
| **Param.** | **Definition** | **I** | **II** | **III** | **IV** | **V** | **VI** |
| $\rho$ | Precision of learning | 0.39 | 0.42 | 0.42 | 0.40 | 0.42 | 0.41 |
| $\rho_0$ | High school signal | 0.49 | 0.43 | 0.43 | 0.51 | 0.51 | 0.47 |
| $\beta$ | Transferability | 0.90 | 0.66 | 0.66 | 0.74 | 0.66 | 0.66 |
| $P$ | Matching premium | 0.20 | 0.155 | 0.155 | 0.20 | 0.155 | 0.155 |
| $R_f(\epsilon_f)$ | Return function | 0.235 | 0.225 | 0.225 | 23 | 0.235 | 0.225 |
| $C_\lambda$ | Expected switching penalty | 1.59 | 1.65 | 1.65 | 1.16 | 1.72 | 0.67 |

TABLE 10. Distance-minimizing parameters using alternate criteria - not weighted

| | | Distance criteria | | | | | |
|---|---|---|---|---|---|---|---|
| **Param.** | **Definition** | **I** | **II** | **III** | **IV** | **V** | **VI** |
| $\rho$ | Precision of learning | 0.38 | 0.41 | 0.41 | 0.39 | 0.42 | 0.40 |
| $\rho_0$ | High school signal | 0.45 | 0.41 | 0.41 | 0.49 | 0.51 | 0.45 |
| $\beta$ | Transferability | 0.82 | 0.66 | 0.66 | 0.66 | 0.66 | 0.66 |
| $P$ | Matching premium | 0.18.5 | 0.155 | 0.155 | 0.20 | 0.155 | 0.155 |
| $R_f(\epsilon_f)$ | Return function | 0.235 | 0.235 | 0.23 | 0.23 | 0.235 | 0.225 |
| $C_\lambda$ | Expected switching penalty | 2.10 | 1.72 | 1.43 | 1.03 | 1.72 | 0.67 |

6.1.3. *Ability.* Figure 9 shows the distribution of timing of specialization for graduates who scored in the upper and lower ability quartiles on their college entrance ACT or SAT exam. The quantitative majors we have retained attract relatively high-ability students; approximately 1/3 of the sample falls into the lower two quartiles, and this sparsity makes the specialization patterns between the two groups difficult to compare. Note that this is particularly the case for Science & Math and Engineering, while Business & Economics students are more evenly distributed across ability groups. For the latter, while lower ability students tend to specialize slightly earlier than their higher ability peers, the two curves are overall quite similar.

Table 11 presents two sets of parameter estimates: our primary estimates for comparison alongside values that best fit the subsample of individuals in the higher




Lower ability Upper ability

Figure 9: Timing of specialization: lower and upper SAT/ACT quartiles

ability bracket. Note that we do not report parameter estimates for the lower ability sample: while our entire sample contains roughly equal shares of students in each major (slanted towards science, the largest category), the subsample of lower-ability graduates is heavily dominated by Business & Economics majors. The high-ability subsample estimates are very similar to the full sample with respect to learning; transferability of education and the matching premium are both a little lower.

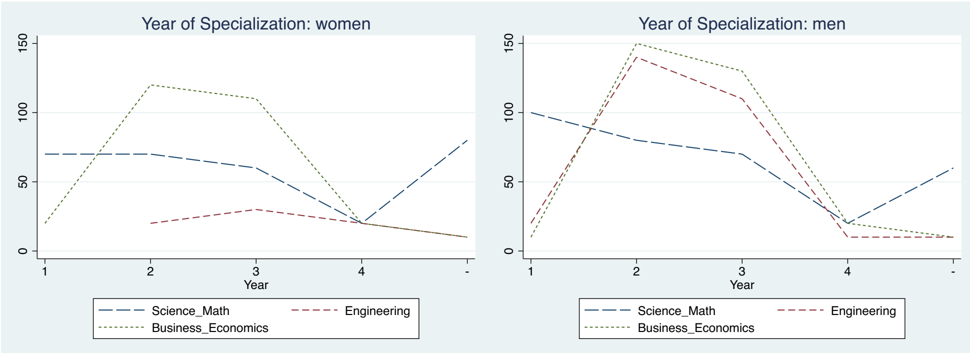
Table 11. Restricted sample: students from upper SAT/ACT quartiles

| Parameter | Definition | Baseline | Upper Qts |
|---|---|---|---|
| $\rho$ | Precision of learning | 0.39 | 0.41 |
| $\rho_0$ | High school signal | 0.49 | 0.49 |
| $\beta$ | Transferability | 0.90 | 0.70 |
| $P$ | Matching premium | 0.20 | 0.17 |
| $R_f(\epsilon_f)$ | Return function | 0.235 | 0.235 |
| $C_\lambda$ | Expected switching penalty | 1.59 | 3.3 |

6.1.4. *Gender.* Does the learning value or transferability of education depend on gender? There is considerable evidence that major choice itself varies across genders.[41] Furthermore, gender-correlated differences in expected labor market attachment could influence the importance of specialized skills vs. information about ones comparative advantage. Bronson (2014) highlights differential penalties in labor supply reductions as one reason women avoid high-paying majors.[42]

[41]See Montmarquette et al. (2002), Kirkeboen (2012), Holzer and Dunlop (2013), Turner and Bowen (1999), Dickson (2010)

[42]Walker and Zhu (2011) also find that returns to majors vary across genders.




Women Men

Figure 10: Timing of specialization: women and men

Gender-correlated differences in risk aversion (De Paola and Gioia (2011)) and competitiveness and overconfidence (Reuben et al. (2013)) can also play a role.

Figure 10 shows the timing of specialization and major choice for men and women. As above, we do not report parameter estimates for women due to the very small share of women majoring in Engineering. A comparison of parameter estimates for men vs. the full sample (see Table 12) shows slightly greater divergence than the ability subsample: learning is faster – both in high school and in mixed-discipline studies – while transferability of education and the matching premium are lower. Importantly, however, the tradeoff of interest remains pertinent: with low transferability of education, field-related education earns a large premium; however, multi-disciplinary studies are informative.

Table 12. Restricted sample: men only

| **Parameter** | **Definition** | **Baseline** | **Men** |
|---|---|---|---|
| $\rho$ | Precision of learning | 0.39 | 0.40 |
| $\rho_0$ | High school signal | 0.49 | 0.52 |
| $\beta$ | Transferability | 0.90 | 0.86 |
| $P$ | Matching premium | 0.20 | 0.238 |
| $R_f(\epsilon_f)$ | Return function | 0.235 | 0.235 |
| $C_\lambda$ | Expected switching penalty | 1.59 | 1.63 |

## 6.2. **Extensions.**

6.2.1. *Relation to Mincer equation specifications.* Our specification, along with the assumption of logarithmic utility, implies that log earnings are a concave function of years of schooling. Assume that the mapping from earnings to flow payoffs is logarithmic and write $w_f = \exp y_f$ for earnings:

$$w_f = \exp\{(R(\epsilon_f)) + \mathbb{1}_{\theta=f}P\} \tag{26}$$

$$\approx \exp\{(R(\epsilon_f))\}(1 + \mathbb{1}_{\theta=f}P) \tag{27}$$

This justifies our interpretation of $P$ as a proportional earnings premium for type-matched agents. When agents apply a logarithmic utility mapping to earnings, we recover the specification in (12).

Equation (12) relates to the *schooling* component of a Mincer equation.[43] Two important features of our specification are at odds with Mincer equations:[44] while Mincer equations use years of schooling as a covariate, we use the effective stock of skills $\epsilon$. Second, a standard Mincer equation has the logarithm of income depend linearly on years of schooling. For comparability, we can use the best linear approximation (in the sense of minimizing quadratic distance) to our estimated returns function, using $\epsilon$ as the covariate, which leads us to retain the value:

$$R(\epsilon_f) = 0.01 + 0.051\epsilon_f. \tag{28}$$

Ignoring informational benefits and field switches, an additional year of schooling increases log earnings by 0.051, corresponding to a 5.1% increase in earnings.

This estimate can be refined in light of our results: taking into account imperfect transferability, $\beta \times 5.1\%$ is a lower bound on the return to schooling. Since about one quarter of agents end up switching fields, $(1/4\beta + 3/4) \times 5.1\%$ gives us an estimate of the average return to specialized schooling. The informational benefits (which increase the probability that a premium will be earned) imply that these are underestimates of the total return to education.

6.2.2. *Overeducation.* While not the focus of this study, our model of higher education has implications for over-education.[45] According to the model, those who choose to work in a field unrelated to their studies will have a smaller stock of specialized education than their fellow graduates who did not change fields. While the

[43]Since our sample contains students of the same age, all of whom attain a bachelor's degree, there is little observable variability in experience, and no observable employment record that would enable tenure observations. Accordingly, our theoretical specification omits experience and tenure effects, leaving only the years of schooling component.
[44]See, for example, Heckman et al. (2006).
[45]See McGuinness (2006) for a review.

allocation of tasks within an occupation grouping is outside our model, it is natural to suppose these individuals will be hired into less-advanced posts than their peers who majored in the occupation-related field. In keeping with Kim et al. (2012)'s study of Korean college graduates, we therefore anticipate a positive correlation between horizontal and vertical occupation-education mis-match.

To investigate this, we take advantage of an additional variable in the data: the (self-reported) education level required by the respondent's most recent occupation. We recode these responses into a binary over-education variable, equal to 1 if the occupation requires less than a bachelor's degree, or if the occupation requires a bachelor's degree and the respondent has earned a master's degree or more.

Table 13 presents results of a regression of occupation-education match on over-education. As expected, we find a positive relationship between over-education and horizontal education-occupation mismatch. Controlling for field of study and occupation, we find that mis-matched workers are approximately 11% more likely to be overeducated. The effect is stronger when restricting to non-quantitative majors: students graduating in these fields are 18% more likely to be overeducated if they have switched to an occupation unrelated to their field of study.

Table 13. Probability of overeducation

| | **Probability of overeducation** | | |
|---|---|---|---|
| | All majors | Quantitative | Non-quantitative |
| Match | -0.108*** | -0.0967*** | -0.183** |
| | (0.000) | (0.004) | (0.050) |
| Controls | X | X | X |
| $R^2$ | 0.097 | 0.100 | 0.107 |
| adj. $R^2$ | 0.089 | 0.092 | 0.083 |
| Sample size | 2110 | 1560 | 550 |

Source: B&B93:03, sample restrictions described in section A.1. P-values in parentheses; ** $p < 0.05$, *** $p < 0.01$. Sample sizes rounded to the nearest 10. Match is a dummy variable equal to 1 if the field of studies is the same as the field of work. Controls are major and occupation dummies.

6.2.3. *Early career labor market rigidities.* While our estimated parameter values are specific to the context under investigation – American college students majoring in one of three quantitative, applied fields – the question we address is not. How would an education system characterized by flexible specialization times perform in different countries? One important mediating factor is the flexibility of the domestic labor market, particularly with respect to early career occupation changes. While we have not modeled the labor market explicitly, the one-time cost incurred by workers who change fields reflects the stickiness of occupation categories, above and beyond the transferability of skills.

Suppose the education system we have modeled was adopted by a country with different labor market conditions. How would the probability of changing fields, and of comparative advantage-occupation mismatch, adjust? To explore these questions, we carry out two experiments. Starting from our baseline parameter estimates, we vary the expected cost of switching fields ($C_\lambda$). Maintaining all other parameters at their estimated levels, we then compute counter-factual labor market outcomes.

Table 14 presents the result of this experiment, with the expected cost of switching set at the high end to 2.02 years of income, and at the low end to 1.17 years of income.[46] As expected, a higher cost of switching fields reduces the probability of changing fields, and decreases the probability of working in the field of comparative advantage. Reducing the expected switching cost produces a symmetric effect. Interestingly, the proportion of students who specialize in the field of comparative advantage is not affected:[47] this suggests that students do not significantly adapt their timing of specialization in light of a change in expected switching costs, but they do adjust their occupation choices.

Table 14. Policy Experiment

| | E(cost) | Initially matched | Change fields | Remain type-mismatched |
|---|---|---|---|---|
| **Baseline** | **1.52** | **53%** | **24%** | **23%** |
| High cost | 2.02 | 53% | 20% | 27% |
| Low cost | 1.17 | 53% | 29% | 18% |

## 7. Conclusion

Does a broad education help people orient themselves towards occupations which are well-suited to them? In the case of post-secondary education, we find evidence that it does. The parameter values we estimate are consistent with a genuine exploration-exploitation tradeoff: broad studies provide information, but

[46]These values are chosen as they represent one step up and one step down on our parameter grid, and are roughly symmetrical increments around our benchmark value.

[47]There is in fact a very small effect, which is not robust to rounding.

specialized studies are more valuable on-the-job. Furthermore, when given the freedom of choice students choose to their timing of specialization in a way consistent with optimal stopping behavior.

While the parameter values we estimate lend support to our model, they also highlight features of the economic environment which are often overlooked. First of all, having explicitly modelled individual heterogeneity as a comparative advantage, we estimate the importance this has on the labor market. The return to working in a field related to one's comparative advantage is large: our estimates put it at 20% of total wages. Secondly, we unpack the college premium along a new dimension: controlling for degree and major, does the timing of specialization matter? The fact that it does suggests that the degree of specialization in college, along with the individual heterogeneity, should be accounted for more carefully when calculating returns to higher education.

Our results point to the importance that education policy has in shaping the labor market returns to education. Since we have not modelled education provision, it is beyond the scope of this paper to assess whether the resulting welfare loss is efficient. However, as the policy experiment in Section 5 illustrates, imposing early specialization is costly to students – and that cost is largely hidden from view.

## References

Altonji, J. G. (1993): "The Demand for and Return to Education When Education Outcomes are Uncertain," *Journal of Labor Economics*, 11, 1–37.

——— (1995): "The effects of high school curriculum on education and labor market outcomes," *The Journal of Human Resources*, 30, 409–438.

Altonji, J. G., E. Blom, and C. Meghir (2012): "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annual Review of Economics*, 4, 185–223.

Altonji, J. G., L. B. Kahn, and J. D. Speer (2013): "Cashier or consultant? Entry labor market conditions, field of study, and career success," working paper.

Arcidiacono, P. (2004): "Ability sorting and the returns to college major," *Journal of Econometrics*, 121, 343–375.

Arcidiacono, P., E. Aucejo, A. Maurel, and T. Ransom (2013a): "College attrition and the dynamics of information revelation," working paper.

Arcidiacono, P., E. M. Aucejo, and V. J. Hotz (2013b): "University differences in the graduation of minorities in STEM fields: Evidence from California," working paper.

Arcidiacono, P., E. M. Aucejo, and K. Spenner (2012): "What Happens After Enrollment? An Analysis of the Time Path of Racial Differences in GPA and Major Choice," *IZA Journal of Labor Economics*, 1(1), 1–24.

Badillo-Amador, L., A. Garca-Snchez, and L. Vila (2005): "Mismatches in the Spanish Labor Market: Education vs. Competence Match," *International Advances in Economic Research*, 11, 93–109.

Beffy, M., D. Fougère, and A. Maurel (2012): "Choosing the field of study in postsecondary education: Do expected earnings matter?" *Review of Economics and Statistics*, 94, 334–347.

Bender, K. A. and J. S. Heywood (2011): "Educational Mismatch and the Careers of Scientists," *Education Economics*, 19, 253–274.

Blom, E. (2012): "Labor market determinants of college major," working paper.

Bordon, P. and C. Fu (2013): "College-Major Choice to College-Then-Major Choice," working paper.

Borghans, L. and B. H. H. Golsteyn (2007): "Skill transferability, regret and mobility," *Applied Economics*, 39, 1663–1677.

Bradley, E. S. (2012): "The Effect of the Business Cycle on Freshman Major Choice," working paper.

Bronson, M. A. (2014): "Degrees Are Forever: Marriage, Educational Investment, and Lifecycle Labor Decisions of Men and Women," working paper.

Cabrales, A., O. Gossner, and R. Serrano (2013): "Entropy and the Value of Information for Investors," *American Economic Review*, 103, 360–77.

Carnevale, A. P., B. Cheah, and J. Strohl (2012): "Hard Times, College Majors, Unemployment and Earnings: Not All College Degrees Are Created Equal." *Georgetown University Center on Education and the Workforce*.

Chevalier, A. (2012): "To Be or Not to Be... a Scientist?" working paper.

De Paola, M. and F. Gioia (2011): "Risk Aversion And Major Choice: Evidence From Italian Students," working paper, University of Calabria.

Dickson, L. (2010): "Race and gender differences in college major choice," *The Annals of the American Academy of Political and Social Science*, 627, 108–124.

Dolton, P. and A. Vignoles (2002): "Is a broader curriculum better?" *Economics of Education Review*, 21, 415–429.

Eeckhout, J. and X. Weng (2011): "Assortative Learning," working paper, Department of Economics, University of Pennsylvania.

Felli, L. and C. Harris (1996): "Learning, Wage Dynamics, and Firm-Specific Human Capital," *Journal of Political Economy*, 104, 838–68.

Finnie, R. (2001): "Graduates' earnings and the job skills-education match," *Education Quarterly Review: Statistics Canada*, 7, 7–21.

——— (2002): "Early Labour Market Outcomes of Recent Canadian University Graduates by Discipline: A Longitudinal, Cross-Cohort Analysis," Statistics Canada Analytical Studies Branch.

Heckman, J., L. Lochner, and P. Todd (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek and F. Welch, Elsevier.

Heijke, H. and C. Meng (2006): "The effects of higher education programme characteristics on allocation and performance of the graduates: A European view," working paper.

Holzer, H. J. and E. Dunlop (2013): "Just the Facts, Ma'am: Postsecondary Education and Labor Market Outcomes in the US," working paper.

Joensen, J. S. and H. S. Nielsen (2009): "Is there a Causal Effect of High School Math on Labor Market Outcomes?" *Journal of Human Resources*, 44, 171–198.

Kim, H.-K., S. C. Ahn, and J. Kim (2012): "Vertical and Horizontal Education-Job Mismatches in the Korean Youth Labor Market : A Quantile Regression Approach," working paper 1201, Research Institute for Market Economy, Sogang University.

Kinsler, J. and R. Pavan (forthcoming): "The Specificity of General Human Capital: Evidence from College Major Choice," *Journal of Labor Economics*.

Kirkeboen, L. J. (2012): "Preferences for lifetime earnings, earnings risk and nonpecuniary attributes in choice of higher education," working paper.

Levine, P. B. and D. J. Zimmerman (1995): "The benefit of additional high-school math and science classes for young men and women," *Journal of Business & Economic Statistics*, 13, 137–149.

Main, J. and B. Ost (2014): "The Impact of Letter Grades on Student Course Selection and Major Choice: Evidence from a Regression-Discontinuity Design," *The Journal of Economic Education*, 1, 1–10.

Malamud, O. (2010): "Breadth versus Depth: The Timing of Specialization in Higher Education," *Labour*, 24, 359–390.

——— (2011): "Discovering One's Talent: Learning from Academic Specialization," *Industrial & Labor Relations Review*, 64, 375–405.

McGuinness, S. (2006): "Overeducation in the Labour Market," *Journal of Economic Surveys*, 20, 387–418.

McGuinness, S. and P. J. Sloane (2011): "Labour market mismatch among UK graduates: An analysis using REFLEX data," *Economics of Education Review*, 30, 130–145.

Montmarquette, C., K. Cannings, and S. Mahseredjian (2002): "How do young people choose college majors?" *Economics of Education Review*, 2002, 543–556.

Morin, L.-P. (2013): "Estimating the benefit of high school for universitybound students: evidence of subjectspecific human capital accumulation," *Canadian Journal of Economics/Revue canadienne d'économique*, 46, 441–468.

Moscarini, G. and L. Smith (2001): "The Optimal Level of Experimentation," *Econometrica*, 69, pp. 1629–1644.

Ost, B. (2010): "The role of peers and grades in determining major persistence in the sciences," *Economics of Education Review*, 29, 923–934.

Papageorgiou, T. (2014): "Worker Sorting and Agglomeration Economies (Online Appendix)," working paper.

Reuben, E., M. Wiswall, and B. Zafar (2013): "Preferences and biases in educational choices and labor market expectations: shrinking the black box of gender," working paper.

Robst, J. (2007): "Education and job match: The relatedness of college major and work," *Economics of Education Review*, 26, 397–407.

Saniter, N. and T. Siedler (2014): "Door Opener or Waste of Time? The Effects of Student Internships on Labor Market Outcomes," working paper.

Silos, P. and E. Smith (forthcoming): "Human capital portfolios," *Review of Economic Dynamics*.

Stange, K. (2013): "Differential Pricing in Undergraduate Education: Effects on Degree Production by Field," working paper.

Stinebrickner, R. and T. Stinebrickner (2014): "A major in science? Initial beliefs and final outcomes for college major and dropout," *The Review of Economic Studies*, 81, 426–472.

Stinebrickner, T. R. and R. Stinebrickner (forthcoming): "Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model," *Journal of Labor Economics*.

Trachter, N. (forthcoming): "Stepping Stone and Option Value in a Model of Postsecondary Education," *Quantitative Economics*.

Turner, S. E. and W. G. Bowen (1999): "Choice of Major: The Changing (Unchanging) Gender Gap," *Industrial & Labor Relations Review*, 52, 289–313.

Walker, I. and Y. Zhu (2011): "Differences by degree: Evidence of the net financial rates of return to undergraduate study for England and Wales," *Economics of Education Review*, 30, 1177–1186.

Wine, J. S., M. B. Caminole, S. Wheeless, K. Dudley, J. Franklin, and K. Perry (2005): "1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03)," *National Center for Education Statistics*, 1–882.

Zafar, B. (2011): "How do college students form expectations?" *Journal of Labor Economics*, 29, 301–348.

## Appendix A. Data

A.1. **Sample.** The B&B 93:03 sample is based on the 1993 National Postsecondary Student Aid Study,[48] restricted to students who were identified as baccalaureate recipients in the 1992-1993 school year. We use the restricted-access version of this dataset to generate the moments we match in our simulation. Since the simulation exercise was not done within the secure data environment, all data used in the estimation – as well as descriptive statistics reported elsewhere in the paper – have to meet disclosure restrictions. This requirement imposes two substantial limitations for our purposes: all sample sizes and frequency counts must be rounded to the nearest 10, and no values can be reported for cells with fewer than 3 individuals. While the estimation procedure uses percentage frequencies rather than count data, the second restriction results in some data loss: this is particularly the case for wage data, which we measure separately for those who switch fields and those who do not, within each major-timing of specialization cell.

While college graduates are likely to be more homogenous in ability than other groups (for instance, college entrants), they are nevertheless a disparate collection of individuals. We restrict the sample in a number of ways, both due to data quality and for conceptual reasons. The entire sample includes 10,980 individuals. We first restrict to individuals who are between 21 and 23 years of age at college graduation, reducing the sample to 7,090. Many of these students transferred institutions some years into their studies. Unfortunately, detailed course data is only available from the degree-granting university; in many cases, transferred courses are noted on the final transcripts, but without a date and often without a course-specific credit value. We retain as many transfer students as we can: essentially, those who transferred after one year or less, and whose transferred courses are identified.[49] This remains a costly restriction, reducing the sample to

[48]The NPSAS is a nationally representative sample of students (and institutions) at all levels of post-secondary education, at all types of institutions.

[49]We do not observe how long a student spent at a different institution. In practice, we allow students to have up to 45 transfer credits, if these transfer credits are or can be associated to a list of transferred courses. Students with more than 45 transfer credits are dropped. Students with between 20 and 45 transfer credits are retained: when applicable, these students are assumed to have 1 year of transferred courses. If a student has more than 20 transfer credits which are not associated to a list of transferred courses, that student is dropped. Students with 20 transfer credits or less are considered to not have taken an additional year, but to have earned these credits in other ways. These transfer credits (whether attributed or otherwise) are coded as part of their first year of studies.

5,750. Cleaning credit values and removing individuals with excessively high or low credit counts, as well as those with study gaps of a year or more, reduces the sample to 5,260. Finally, we restrict to those individuals who were followed up in 2003, leaving a sample of 4,170.[50]

A.1.1. *Majors.* While college may prepare students for work in many different ways, our model is constructed with applied majors in mind. The tradeoff between between skills and information only has bite when students anticipate that additional course material in their eventual field of work will bring additional returns. Many majors have weak links to the labor market; while this does not mean that they do not yield returns, optimal course choices in such majors may very well follow a different path from those in applied fields. We therefore restrict our sample to students with applied majors: these include six fields and 2160 individuals, subdivided into quantitative fields (1580) and non-quantitative fields (580).[51] While our estimation is done using only the quantitative graduates, we present statistics for all six majors here. The allocation across majors is given in Table 15.

TABLE 15. Count of observations by major

| **Major** | Count | *Quantitative* | *Non-quantitative* |
|---|---|---|---|
| Science | 630 | X | |
| Engineering | 350 | X | |
| Business & Econ | 600 | X | |
| Education | 430 | | X |
| Nursing | 70 | | X |
| Social Wk & Protective | 80 | | X |
| Total | 2160 | 1580 | 580 |

Source: B&B93:03, sample restrictions described in section A.1. Counts rounded to the nearest 10 to respect disclosure restrictions.

Our analysis abstracts from the vertical dimension of ability. Computational constraints require us to be parsimonious with parameters, and we are specifically interested in horizontal abilities. Restricting our analysis to individuals who

[50]In our analysis we make use of the 2003 occupation observation only. This refers to the occupation held most recently by the respondent at the time of the survey, and therefore is non-missing even for individuals who are unemployed at the time of the survey. For most individuals, however, this is the occupation held ten years after college graduation.

[51]While this figure represents a dramatic reduction of the original sample, it is worth noting that this is largely due to the vast heterogeneity in the college graduate population. Using a sister dataset, albeit with an even more heterogenous population, Silos and Smith (forthcoming) are required to make similarly harsh restrictions.

earned a bachelor degree between ages 21 and 23 narrows the distribution of ability within the sample. We remain concerned, however, that some skills may act as gatekeepers to certain fields, preventing lower-ability students from completing majors in those subjects even had they wished to.

Figure 11 shows the distribution of individuals across quartiles of SAT or ACT scores, by major. Clearly, some majors have a greater mass of high-ability students than others. This, combined with the small number of individuals choosing nursing and social work majors, motivates us to use only the three quantitative majors for our primary specification. While this restricts the interpretation of our results, it is plausible that this set of students is more homogenous that those enrolled in all six fields combined.

A.1.2. *Term length.* In principle, school terms are an intuitive and straight-forward concept, and relate naturally to the discrete-time version of our model. Classes are chosen at the start of the term and difficult to adjust once the term is underway; at the end of each term, enrollment for the next term – and the associated course selection – gets underway. In practice, however, the concept of a school term is difficult to pin down. In addition to the diversity of term structures (semesters, trimester, quarters, etc.), there are many students who enter university with some number of college credits. These may have been earned by exam (for instance, Advanced Placement courses), taken while in high school, or earned at a previous university and transferred to the degree-granting institution.[52] Even defining the start of the school year is not without difficulties: fall courses at one institution may start before summer courses have completed at another.

To mitigate these problems, we use the academic year as our period length; the remaining issues are dealt with in one of two ways. First, having restricted our sample to individuals who graduate between the ages of 21 and 23 – and therefore eliminating students who take an unusually long time to complete their degree – we abstract from term dates and divide a student's courses chronologically into four terms of equal credit value.

[52]In general, credits earned through exam are indicated as such on the transcript. Given that these are not actually classes, and are often earned based on prior education, they are not included in the analysis.

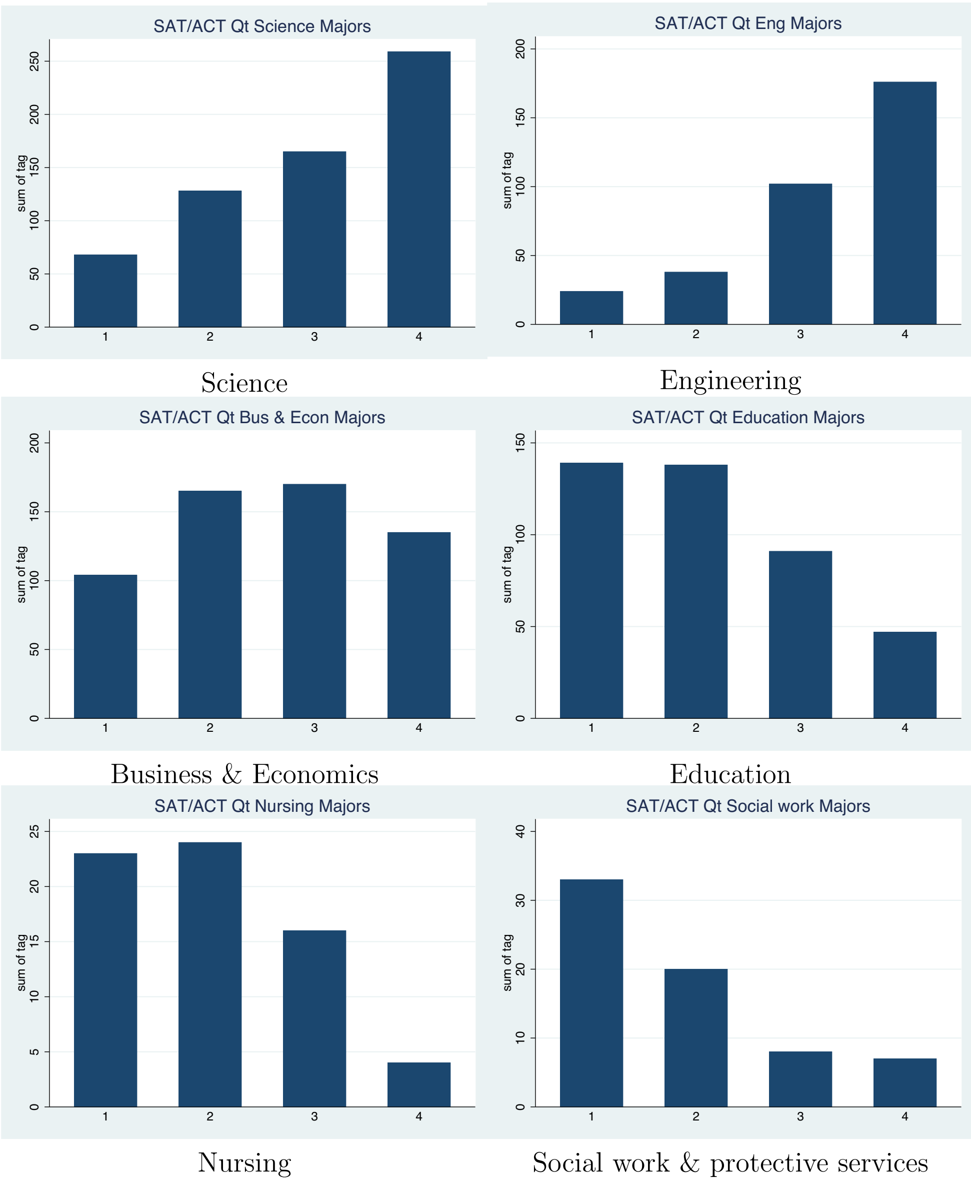



Figure 11: SAT or ACT quartiles, by major

Our second approach accommodates diversity in time-to-degree by coding school years as faithfully as possible. We define the academic year as running from August to July, and attribute courses accordingly. To avoid creating spurious years of college (due to a late summer course, for instance), we recode any terms with 6 or fewer credits as belonging to the next or previous school term. Finally, to maintain a reasonably homogenous group, we restrict the sample to students who graduate in 4 or 5 years. Table 16 gives the distribution of time-to-degree for these students.

Table 16. Time to degree, by major

| **Major** | 4 years | 5 years |
|---|---|---|
| Science | 380 | 210 |
| Engineering | 140 | 170 |
| Business & Econ | 350 | 210 |
| Education | 200 | 180 |
| Nursing | 40 | 30 |
| Social Wk & Protective | 40 | 30 |
| Total | 1150 | 830 |

Source: B&B93:03, sample restrictions described in section A.1. The timing of specialization used is the primary specification, with time-to-degree computed using a true-term approach.

Table 17 presents correlations between two different timing variables computed using both true years and standardized years. The two approaches are strongly correlated. While the second approach would permit us to examine how the timing of specialization is related to time-to-degree, data sparsity becomes a pressing concern (note that this would require us to track of 4- and 5-year degree students separately). In addition, it is not clear how adequately our model captures the choice to pursue a 5th year of undergraduate study. For these reasons we use the first approach for our primary specification, standardizing the duration of a college degree to four years.

Table 17. Correlation between timing of specialization using true and standardized years

| Timing variable | TY-con | TY-ret | 4Y-con | 4Y-ret |
|---|---|---|---|---|
| True years - concentration | 1.0000 | | | |
| True years - 90% retention | 0.8277* | 1.0000 | | |
| Four years - concentration | 0.8288* | 0.8190* | 1.0000 | |
| Four years - 90% retention | 0.7512* | 0.8298* | 0.9009* | 1.0000 |

Source: B&B93:03, sample restrictions described in section A.1. Star indicates significance at the 10% level. Thresholds are defined and explained in Section A.3.1.

A.2. **Matching occupations to majors.** One of the key outcomes we are interested in is whether individuals pursue a career in their field of studies, or whether they switch into a different field. This link is better defined for some fields than for others: some majors, such as engineering or education, have obvious careers associated to them. Other majors, including most of the humanities and social sciences, do not lead unambiguously to a certain occupation. Table 18 gives an overview of the occupations held by sample members, as a percentage of all graduates from each major.

Table 18. Share of major in each occupation

| **Occupation** | Sci | Eng | Bus | Edu | Nur | Swp |
|---|---|---|---|---|---|---|
| Educators | 14 | 3 | 5 | 74 | 0 | 14 |
| Business/management | 14 | 24 | 57 | 7 | 0 | 14 |
| Engineering/Architecture | 5 | 44 | 2 | 0 | 0 | 0 |
| Computer science | 11 | 9 | 5 | 0 | 0 | 0 |
| Medical professions | 27 | 0 | 2 | 2 | 100 | 14 |
| Editors/writers/performers | 2 | 0 | 2 | 2 | 0 | 0 |
| Human/protective services/legal pro | 3 | 3 | 7 | 2 | 0 | 43 |
| Research/scientists/technical | 14 | 9 | 2 | 2 | 0 | 0 |
| Administrative/clerical/legal | 2 | 0 | 3 | 2 | 0 | 14 |
| Mechanics/laborers | 2 | 3 | 2 | 2 | 0 | 0 |
| Service industries | 5 | 3 | 13 | 5 | 0 | 0 |
| Other/military | 2 | 3 | 2 | 0 | 0 | 0 |

Source: B&B93:03, sample restrictions described in section A.1. Note that columns may not sum to 100 due to rounding.

Matching occupations with majors is facilitated by the restriction to applied fields described in Appendix A.1. We base our major categories on 14 aggregated majors given in the data,[53] and derive a correspondence between these majors and 12 occupation categories. The aggregation of majors is given in Table 19, along with the corresponding occupations. Note that we select the sample based on major, making no restrictions with regard to occupation.

## A.3. **Timing of specialization.**

A.3.1. *Alternate specifications.* In addition to the primary specification described in Section 2.2, we derive a number of alternate timing of specialization variables. These alternate approaches allow us to run robustness checks and explore the generality of our primary measure.

Figures 12 - 14 present the distribution of timing of specialization for four alternate specifications. Five different thresholds are represented for each specification,

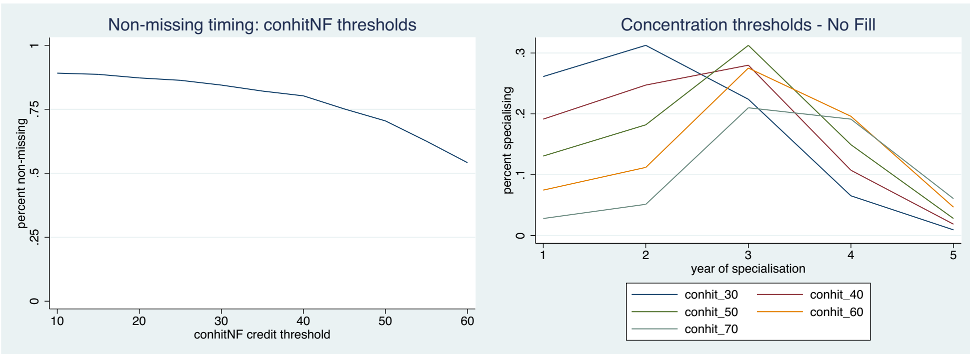
[53]We deviate from these categories by recoding economics with business, rather than with social sciences.

TABLE 19. Occupation-Major Correspondance

| 14 MAJORS | 6 MATCH CATEGORIES | 12 OCCUPATIONS |
|---|---|---|
| Biological/interdisc sciences<br>Mathematics/physical sciences<br>Computer science | Sciences | Research, Science & Technical<br>Medical Professionals<br>Computer science |
| Engineering/architecture | Engineering & Architecture | Engineering/architecture |
| Business<br>Soc sciences (econ only) | Business & Management | Business and management |
| Education | Education | Educators |
| Health/nursing | Nursing | Medical professionals |
| Social work/protective serv | Human/Protective Services & Legal | Human/protective service/legal prof |
| | *Unmatched occupations* | Administrative/clerical/legal sup<br>Mechanics, laborers<br>Service industries<br>Other, military<br>Editors/writers/performers |
| Humanities/liberal studies<br>Agriculture<br>Social sciences (non-econ)<br>Other<br>Communications/journalism<br>Health/other | *Excluded majors* | |

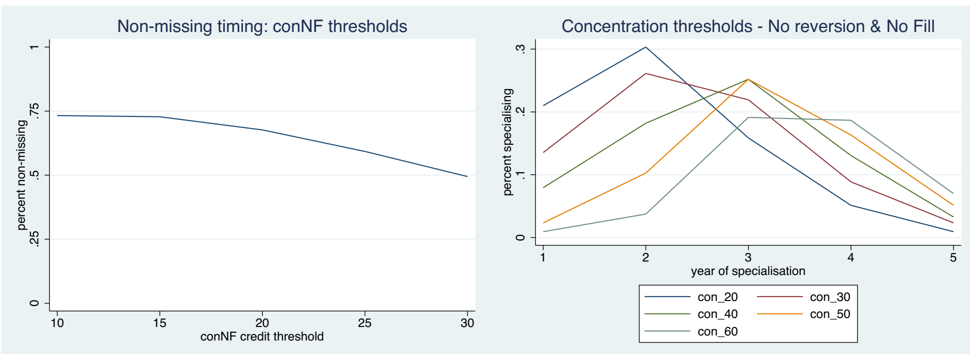
Source: authors' aggregation, based on B&B93:03 major and occupation codes. Note: 'other' majors are not considered matched with 'other' occupations.

as well as the attrition in the number of specializers as the threshold rises. Figures 12 and 13 are computed using true years (see section A.1.2). Both thresholds refer to a concentration of credits the student must reach or exceed; however, the thresholds in Figure 13 require that students remain above that threshold until graduation, while those in Figures 12 do not. Figure 14 presents the distribution of the timing of specialization for equivalent thresholds, using standardized 4-year college tenures. In this figure 'never'-specializers appear on the right as if they had specialized in Year 5 (hence their representation with one graph rather than two).



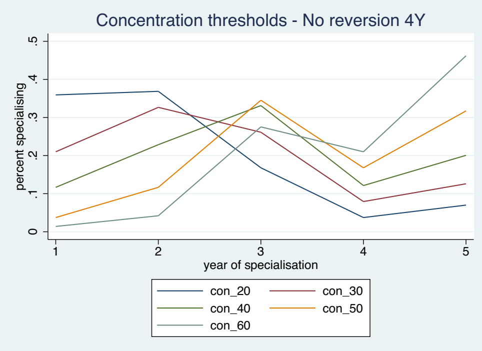

Count of specializers as threshold rises Timing of Specialization

Figure 12: Concentration-hit thresholds, allowing reversion: true years

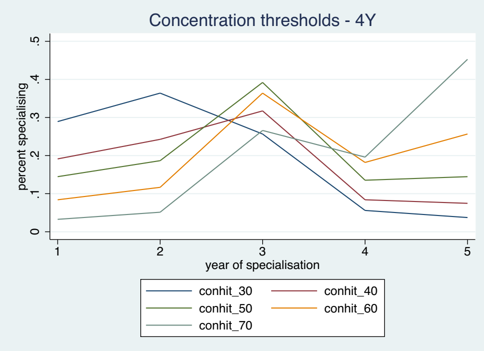



Count of specializers as threshold rises Timing of Specialization

Figure 13: Concentration-hit thresholds, no with reversion: true years




Concentration-hit thresholds, no with reversion




Concentration-hit thresholds, allowing reversion

Figure 14: Timing of specialization using standardized years (note: specialization in Year 5 is equivalent to never specializing)

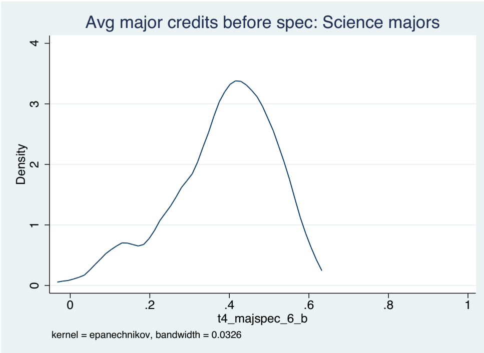
A.3.2. *Specialization and course choices.* We have identified specialization based on the concentration of courses a student takes during each school year. While we have no way to verify our specialization concept externally, we can make some basic checks.

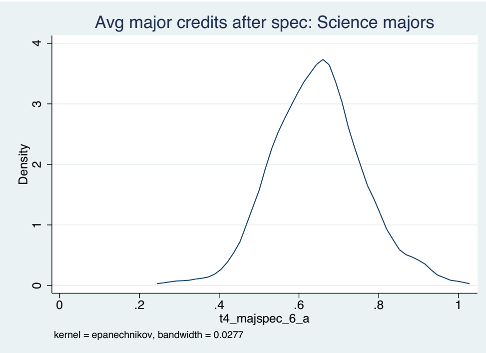
First of all, our approach relies on students taking more major-field courses later in their degrees. Figures 15 and 16 show the distribution of the percent of credits taken in the major field, before and after specialization. While the specification itself could induce a modest difference in these distributions, it is encouraging to see that students are indeed taking few credits in their major prior to specialization.[54]

Next, our model supposes that students take a constant share of courses in their major field in each period of specialized studies, regardless of the timing of specialization. This does not permit late specializers to load up on major-specific courses in order to meet major requirements or catch up with their early-specializing peers. To check whether this assumption is reasonable we look at how total credits and credits in the major field vary with the timing of specialization. Table 20 gives the mean and standard deviation of total credits, for each major and timing of specialization. While there is some variation across majors, the average credit load is encouragingly flat with regards to the timing of specialization.[55]

[54]Science remains an outlier, with the mean credit share before specialization being quite large at 40%, while the mean credit share afterwards is a more typical 60%.

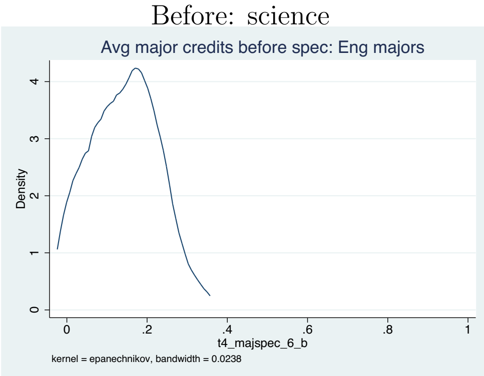
[55]The invariance of total credits to the timing of specialization suggests that the timing of specialization may not be that closely correlated with time-to-degree: late specializers do not

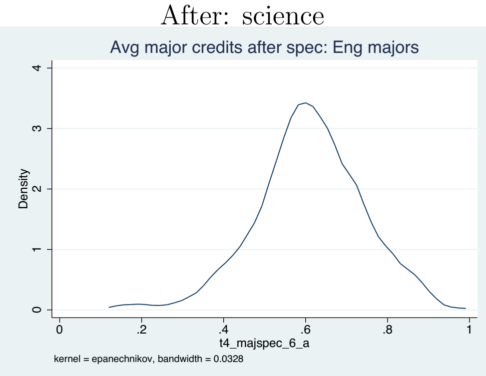



Before: science

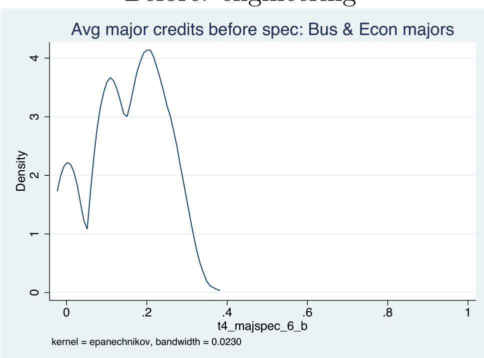
After: science

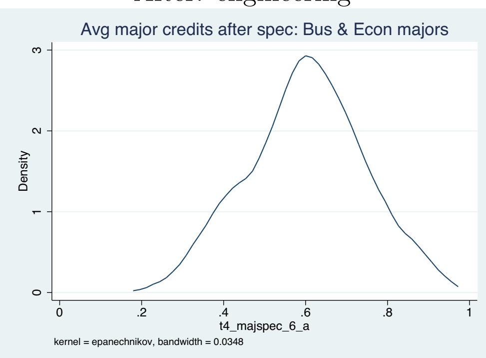



Before: engineering

After: engineering



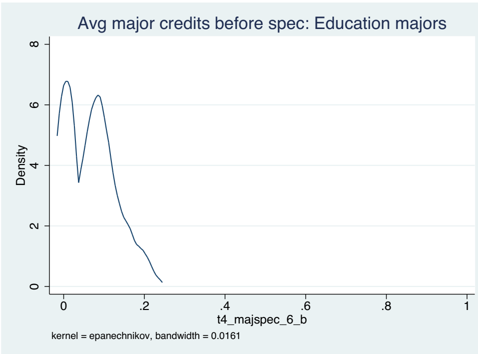

Before: business & econ

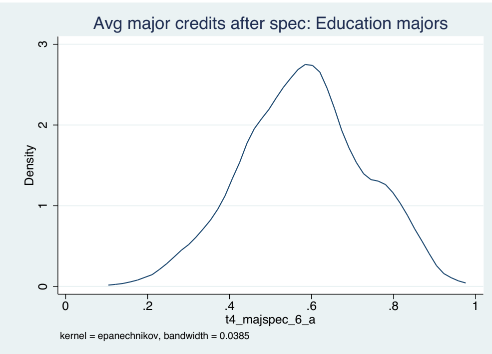
After: business & econ

Figure 15: In-major credit share before and after specialization: quantitative majors (primary specification)

The mean number of in-major credits for each major and timing of specialization are shown in Table 21. Unlike for total credits, the number of in-major credits declines with later specialization, although clearly this is more true in some majors than in others. Late specializers may indeed try to 'catch up' by taking a heavier course load; however, early specializers still take more credits in their major field.

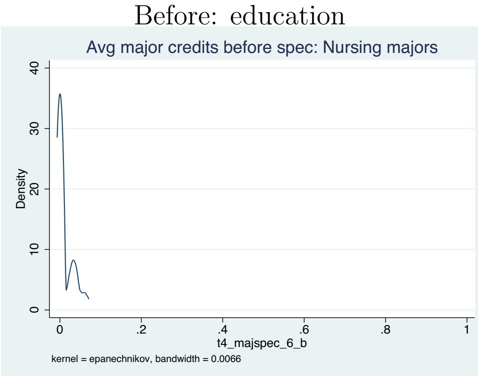
systematically accumulate an extra year of courses. The abstraction we make from time-to-degree is less striking in light of this.

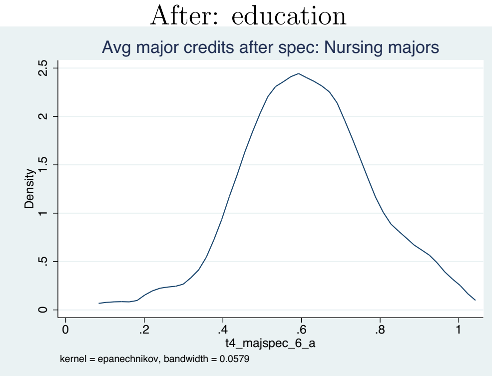



Before: education

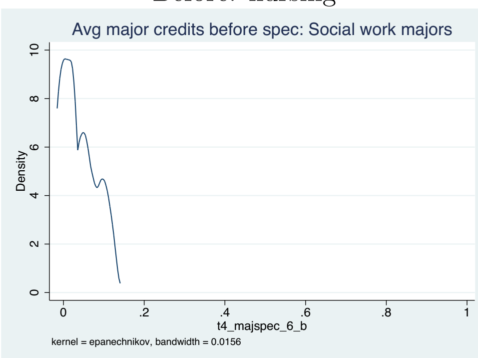
After: education

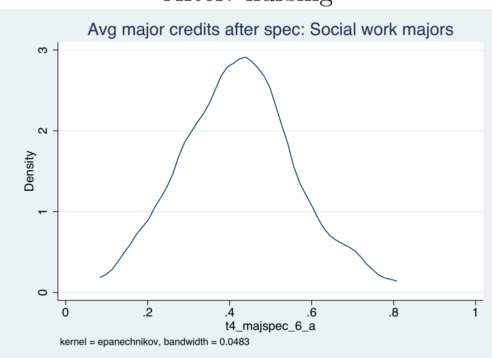



Before: nursing

After: nursing




Before: social work & prot.

After: social work & prot.

Figure 16: In-major credit share before and after specialization: non-quantitative majors (primary specification)

Finally, Table 22 shows the difference in the term-by-term share of courses taken in the major field, before and after specialization. From the table, it does not appear that late specializers experience a higher jump in course taking than early specializers. This suggests that any catch-up by later specializers is modest.

A.4. **Correlations on observables.** Table 23 presents the correlations of our main timing of specialization variable with several observable characteristics of

Table 20. Total credits, by major and timing of specialization

| **Major** | Year 1 | Year 2 | Year 3 | Year 4 | Never |
|---|---|---|---|---|---|
| Science | 126 (10) | 127 (12) | 127 (14) | 129 (15) | 127 (13) |
| Engineering | 134 (12) | 135 (11) | 136 (12) | 132 (17) | 132 (8) |
| Business & Econ | 125 (14) | 126 (11) | 126 (10) | 126 (15) | 125 (12) |
| Education | 133 (13) | 133 (12) | 134 (13) | 136 (15) | 131 (15) |
| Nursing | 125 (7 ) | 128 (15) | 138 (14) | - | - |
| Social Wk & Protective | 129 (11) | 122 (10) | 128 (11) | 126 (12) | - |

Source: B&B93:03, sample restrictions described in section A.1. Standard deviations in parenthesis. A standard undergraduate degree is 120 credits.

Table 21. In-major credits, by major and timing of specialization

| **Major** | Year 1 | Year 2 | Year 3 | Year 4 | Never |
|---|---|---|---|---|---|
| Science | 77 (13) | 74 (12) | 70 (13) | 65 (9) | 39 (18) |
| Engineering | 70 (13) | 63 (12) | 57 (11) | 38 (7) | 10 (9) |
| Business & Econ | 62 (13) | 60 (12) | 52 (12) | 38 (7) | 25 (11) |
| Education | 67 (17) | 60 (14) | 47 (13) | 28 (7) | 8 (6) |
| Nursing | 59 (8) | 55 (7) | 49 (12) | - | - |
| Social Wk & Protective | 41 (13) | 37 (7) | 32 (9 ) | 25 (8) | - |

Source: B&B93:03, sample restrictions described in section A.1. Standard deviations in parenthesis. A standard undergraduate degree is 120 credits.

Table 22. Major-field course share: *after* minus *before* specialization

| **Major** | Year 2 | Year 3 | Year 4 |
|---|---|---|---|
| Science | 0.21 | 0.25 | 0.33 |
| Engineering | 0.46 | 0.5 | 0.33 |
| Business & Econ | 0.48 | 0.45 | 0.34 |
| Education | 0.51 | 0.54 | 0.47 |
| Nursing | 0.56 | 0.68 | - |
| Social Wk & Protective | 0.36 | 0.41 | 0.48 |

Source: B&B93:03, sample restrictions described in section A.1. Standard deviations in parenthesis. A standard undergraduate degree is 120 credits.

the students: family income quartile, father's and mother's education, academic ability prior to college (captured by SAT or ACT score quartiles, and also SAT math and verbal scores separately for those students who took the SAT), and gender. The correlations are small and in general not significant at the 10% level. Family income is the exception: higher incomes are associated with a later timing of specialization.

The absence of correlations may partly be an artifact of aggregation. Table 24 breaks the sample into students graduating with quantitative majors (science, engineering or business and economics), and those graduating with non-quantitative

Table 23. Correlations on Observables - all majors

| | TIM | FaInc | FEdu | MEdu | S/Aq | SATV | SATM | G |
|---|---|---|---|---|---|---|---|---|
| TIMING of SPEC | 1.000 | | | | | | | |
| Family income | 0.049* | 1.000 | | | | | | |
| Father's ed | 0.024 | 0.363* | 1.000 | | | | | |
| Mother's ed | 0.046 | 0.313* | 0.551* | 1.000 | | | | |
| SAT/ACT quart | 0.024 | 0.171* | 0.224* | 0.205* | 1.000 | | | |
| SAT verbal | 0.055 | -0.047 | 0.023 | 0.018 | -0.013 | 1.000 | | |
| SAT math | 0.035 | -0.063* | -0.018 | -0.019 | 0.002 | 0.672* | 1.000 | |
| Gender | -0.020 | -0.065* | -0.070* | -0.038* | -0.178* | -0.007 | -0.037 | 1.000 |

Source: B&B93:03, sample restrictions described in section A.1. Timing of specialization is the primary specification (see Section 2.2). Gender is increasing in femininity. Star indicates significance at the 10% level. Correlations are pairwise, starting from a maximum sample of 2160 observations.

majors (education, nursing or social work and protective services). While the correlations remain modest in size, some stronger patterns emerge. Quantitative graduates are more likely to specialize early if they have higher SAT or ACT scores, while the reverse is true for non-quantitative graduates. On the other hand, non-quantitative graduates are more likely to specialize early if they are women, while the reverse is true in the quantitative fields. The correlation between timing of specialization and family income disappears when considering quantitative majors alone, but strengthens slightly for non-quantitative majors. These correlations motivate our choice of split-sample robustness checks, presented in Section 6. In particular, we consider separately men and women, and upper and lower ability students. Given that the correlation with family income is not present in our primary sample, and is otherwise relatively small, we do not pursue it at this time.

## Appendix B. Analytical results in the linear-return model

B.1. **Linear-return model.** To explore the impact that the parameters of the model have on the empirical outcomes we are interested in, we consider a simplified parametric version of the model presented in Section 3.1. We assume perfect symmetry between the fields ($R_s = R_a$) and a neutral prior, $p_0 = 1/2$. Furthermore, let the return be linear and composed of a baseline wage level and a term proportional to the effective education stock

$$R(\epsilon_f) = w_0 + y\epsilon_f. \tag{29}$$

Table 24. Correlations on Observables - by type of major

| | TIM | FaInc | FEdu | MEdu | S/Aq | SATV | SATM | G |
|---|---|---|---|---|---|---|---|---|
| *Quantitative* | | | | | | | | |
| TIMING of SPEC | 1.000 | | | | | | | |
| Family income | -0.013 | 1.000 | | | | | | |
| Father's ed | 0.015 | 0.340* | 1.000 | | | | | |
| Mother's ed | 0.008 | 0.251* | 0.546* | 1.000 | | | | |
| SAT/ACT quart | -0.104* | 0.131* | 0.197* | 0.192* | 1.000 | | | |
| SAT verbal | 0.045 | -0.108* | -0.041 | -0.006 | -0.049 | 1.000 | | |
| SAT math | 0.017 | -0.055 | -0.100 | -0.037 | 0.007 | 0.653* | 1.000 | |
| Gender | 0.091* | -0.060* | -0.044* | -0.040 | -0.153* | 0.077 | 0.015 | 1.000 |
| *Non-quantitative* | | | | | | | | |
| TIMING of SPEC | 1.000 | | | | | | | |
| Family income | 0.071* | 1.000 | | | | | | |
| Father's ed | 0.038 | 0.301* | 1.000 | | | | | |
| Mother's ed | 0.033 | 0.245* | 0.590* | 1.000 | | | | |
| SAT/ACT quart | 0.101* | 0.110* | 0.166* | 0.155* | 1.000 | | | |
| SAT verbal | 0.107 | 0.041 | 0.094 | -0.134 | 0.048 | 1.000 | | |
| SAT math | 0.088 | -0.001 | 0.035 | -0.080 | -0.063 | 0.657* | 1.000 | |
| Gender | -0.149* | 0.009 | 0.037 | 0.061 | -0.007 | -0.141 | -0.005 | 1.000 |

Source: B&B93:03, sample restrictions described in section A.1. Timing of specialization is the primary specification (see Section 2.2). Gender is increasing in femininity. Star indicates significance at the 10% level. Correlations are pairwise, starting from maximum samples of 1580 (quantitative) and 580 (non-quantitative) observations.

Under these assumptions, we can obtain simple and interpretable closed-form solutions for the optimal length of specialized schooling. Assuming that the agent with education stocks $(e_S, e_A)$ anticipates to remain in field $s$ (even if he learns that his comparative advantage is in field $a$), he will choose the value of $H^*$ in order to achieve an optimal aggregate education stock $\epsilon_S$. The optimal value $H^*$ satisfies:

$$e_S + H^* + \beta e_A = \frac{1}{r} - \frac{w_0 + pP - z}{y_s}, \tag{30}$$

where $z$ is the flow payoff from studies. The agent's optimal tenure in specialized education is $H = H^*$ if positive, and zero otherwise. A necessary condition for $H^*$ to be positive[56] is that inequality (31) hold, which happens when the labor market rewards training, flow payoffs of specialized education are large and the agent is patient.

[56]Inequality (31) obtains from (30) by imposing $e_S = e_A = 0$ and $p = 0$ and imposing that the right-hand side be positive.

$$P + 2w_0 < \frac{2y}{r} + 2z \tag{31}$$

We assume that condition (31) holds strictly. Under these conditions, the value of starting specialized studies in subject $S$ at time $t$ is given by:

$$V_{S,st}(p, t/2, t/2) = r^{-2}\exp\left\{r\frac{(\beta+1)}{2}t + r\frac{pP}{y} + \frac{w_0 r - rz}{y_s} - 1\right\}y + r^{-1}z, \tag{32}$$

where subscript $st$ indicates the intention to stay in field $s$ (as opposed to $sw$ for switching). A similar formula can be obtained if the agent anticipates changing fields. Again assuming an interior choice of $H^*$, we have:

$$V_{S,sw}(p, t/2, t/2) =$$

$$r^{-1}z + r^{-2}y\,(p\,(1-\beta)+\beta)\exp\left\{\frac{r\,(\beta+1)\,t}{2\,(p\,(1-\beta)+\beta)} + \frac{r\,(w_0+P-z)}{y\,(p\,(1-\beta)+\beta)} - 1\right\}. \tag{33}$$

Using these formulas, we can show that some experimentation is worthwhile as long as the signal is informative, thereby guaranteeing $\tau_1 > 0$.

**Proposition 1** (Minimum level of experimentation)**.** *In the linear-symmetric case, $p_s(0)$ is bounded away from $\frac{1}{2}$: the agent engages in mixed studies in the neighborhood of $t = 0$.*

We can further establish that mixed education becomes dominated in finite time: there exists a limit at which agents stop experimenting regardless of the path followed by the belief process.

**Proposition 2** (Maximum level of experimentation)**.** *In the linear-symmetric case, there exists $t^* > 0$ such that $p_s(t^*) = p_a(t^*) = 1/2$: the agent never engages in mixed studies for $t \geq t^*$.*

These two propositions show that the boundary $p_s$ moves from a position bounded away from $1/2$ at $t = 0$ down to $1/2$ before a fixed time $t^*$. While this is not inconsistent with the boundary being locally increasing, it must be decreasing on average. In numerical simulations, the optimal boundary is monotonically decreasing.

## Appendix C. Simulations results in the linear-symmetric model

We illustrate the properties of the optimal solution by identifying the optimal boundary numerically for a given set of parameters in a linear-symmetric model with a deterministic switching cost. We show how parameter changes affect that optimal boundary, the flow of agents through the different education regimes as well as the level of confidence at which their initial specialization takes place. This justifies the parameter selection for our estimation exercise.

C.0.1. *Optimal boundaries and exit.* Figure 17 (reproduced from Section 3.1) displays the optimal belief boundaries, overlaid with a sample belief path of a type $S$ agent. Near $t = 0$, agents require strong beliefs in order to specialize: above 0.75 (to specialize in $S$) or below 0.25 (to specialize in $A$). If they have not specialized by time $t = 8$, a very small deviation of $p_t$ away from $1/2$ is enough to trigger specialization. The density of specialization times is also displayed in the figure, for specialization into $S$ or $A$ separately.[57] By integrating these densities, we find that approximately 68 percent of type $S$ agents eventually specialize correctly in field $S$ – the field of their comparative advantage, while 32 percent specialize in field $A$ and are therefore initially mismatched.

Upon specialization, agents already know the probability with which they will switch fields. The duration of specialized studies is then determined in part by whether the agent will choose to pursue his comparative advantage, should he discover on the labor market that his comparative advantage is in the other field. Figure 18 shows the probability of switching fields as a function of the timing of specialization (solid red line). Note that the earliest specializers do not switch fields. The Figure also displays the length of specialized studies for these same individuals (dashed blue line). Observe that early specializers do not "hedge their bets": they invest heavily in specialized education and plan to stay in their initial field even if their comparative advantage is elsewhere. In the opposite extreme, very late specializers - who have chosen their field on relatively poor information - know that they are likely to discover their comparative advantage is actually in the other field and accordingly spend comparatively little time specializing before

[57] We also assume that the initial distribution of priors is a Dirac mass at $1/2$.

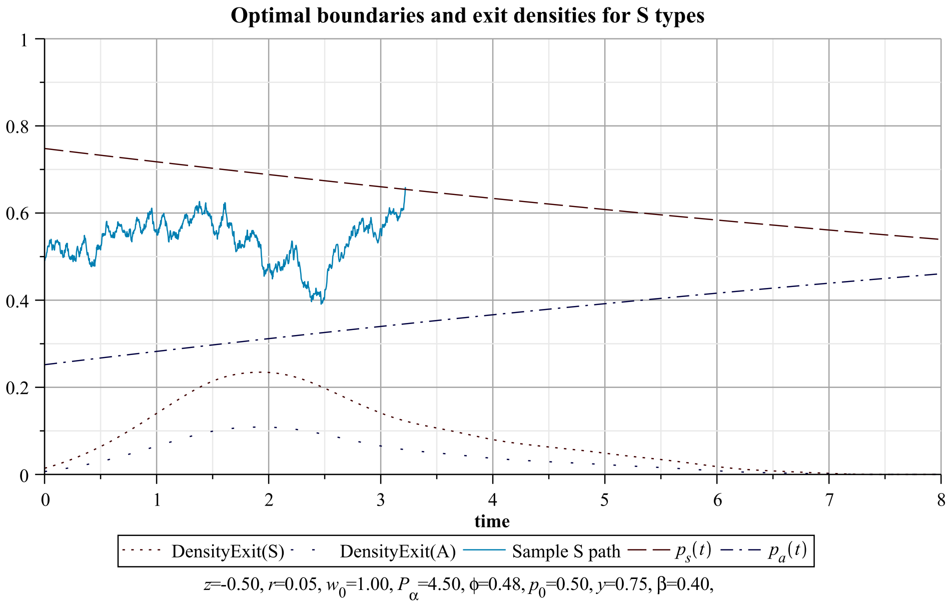



Figure 17: Optimal Boundaries and density of specialization times. *The agent starts specialized education once the belief process $p_t$ escapes the interval* $(p_a(t), p_s(t))$. This particular sample path leads to specialization time $\tau_1 \approx 1.64$ and corresponds to correct specialization (in subject $S$).

joining the labor market (keep in mind that late specializers have accumulated a significant amount of human capital in *both* fields).

C.0.2. *Parameter changes.* We now consider how changes in individual parameter values affect the behavior of agents, *ceteris paribus*. Payoff-relevant parameters $(P, w_0, y, z, \beta, r)$ do not impact the informativeness of signals; however, they influence the relative value of specialized education and type-discovery. Figure 19 illustrates how the belief boundaries change following a discrete change in the value of each of these parameters (the figures show only the upper boundary $p_s(t)$; the lower boundary $p_a(t)$ will adjust symmetrically). Since the speed at which agents reach the boundaries is not affected by these parameters, the location of the boundary itself summarizes the impact of these changes.

While the distributions in Figure 17 are generated under the assumption that all belief paths begin at $p_0 = 1/2$ (and evolve according to equation (2)), the model does allow for agents to have some information about their type at $t = 0$. This is formally represented by a distribution of time-zero beliefs $p_0$ that is correlated with

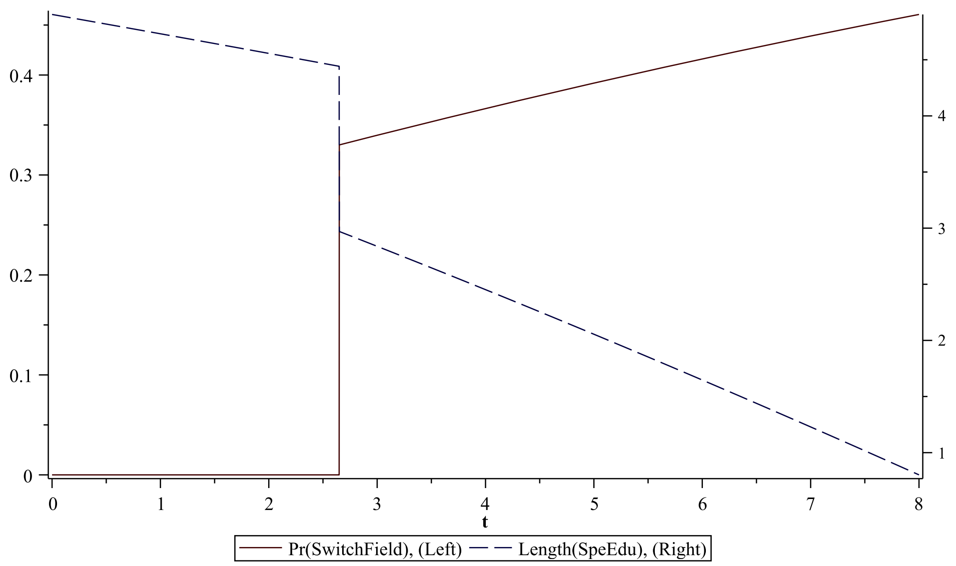
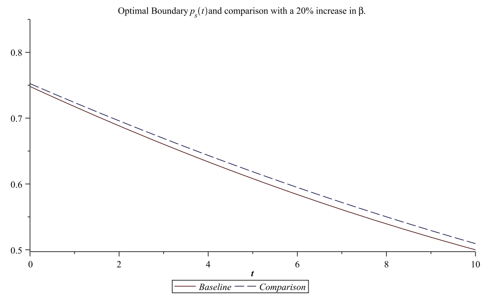



Figure 18: Probability of switching fields and length of specialized education tenure conditional on specialization time. *Early specializers ($t \leq 2.7$) remain in their field of specialization even if revealed to be of type A. They choose a longer education tenure.*

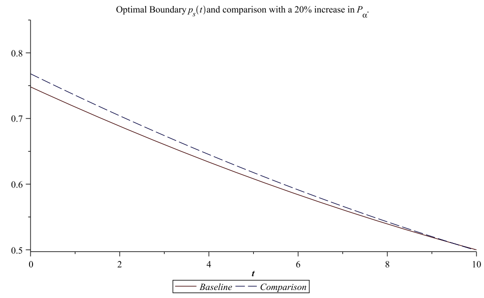
the true type. In contrast with the return parameters, the initial information set does not affect the optimal forward-looking experimentation policy, which implies that its effect is entirely driven by the density of exit times.

Informative initial beliefs are illustrated in Figure 20. Let the belief at $t = 0$ be informative in the following sense: before $t = 0$, agents receive a symmetric binary signal that agrees with their true type with probability $3/5$ and is misleading otherwise.[58] Figure 20 plots the density of timings of specialization – separately for each field – for two populations of $S$-type agents which differ only in their prior information. The solid red line and dashed blue line give the specialization times for agents with an informative prior, while the dotted red line and dot-dashed blue line give the distribution for agents with no prior information about their type. As before, the solid and dotted lines correspond to those agents who specialize – correctly – into field $S$, while the dashed lines represent agents who mistakenly believe their comparative advantage to be in field $A$. We can see that

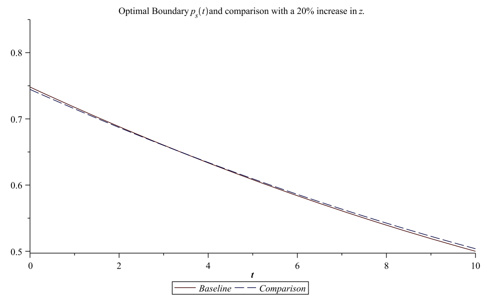
[58]Such a signal leads to type $S$ agents holding updated belief $p_0 = 3/5$ with probability $3/5$ and belief $p_0 = 2/5$ with probability $2/5$, with symmetric numbers for type-A agents.



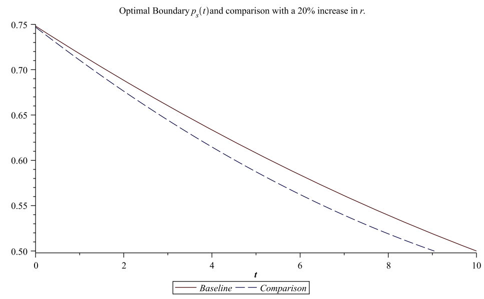

Transferability of human capita: $\beta$



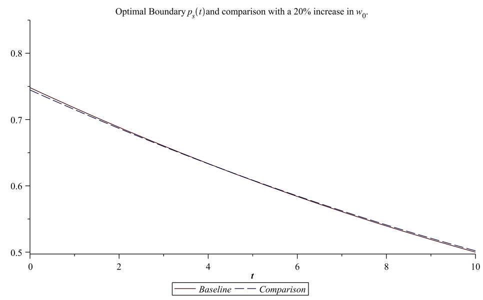

Comparative advantage premium: $P$



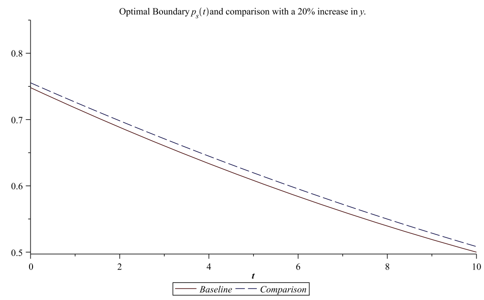

Flow payoffs during studies: $z$




Discount rate: $r$




Returns function intercept: $w_0$




Marginal returns to human capital: $y$

Figure 19: **Comparative statics**. Each sub-figure compares the optimal upper boundary $p_s(t)$ before (solid red line) and after (dashed blue line) a 20% increase of the parameter in question. For a fixed learning technology, an upwards shift of the boundary implies that agents specialize later: they hold out for more confidence before committing to either field.

improving the initial information of agents has the effect of speeding up the process in the sense of first-order stochastic dominance and of increasing the probability of correct initial match.

The precision of learning is the only parameter entering equation (2), hence the beliefs updating process. As a result, it impacts the optimal specialization decision through the choice of boundaries $p_s(t), p_a(t)$, but also the distribution of times at which boundaries are reached and the likelihood of specializing correctly. Figure

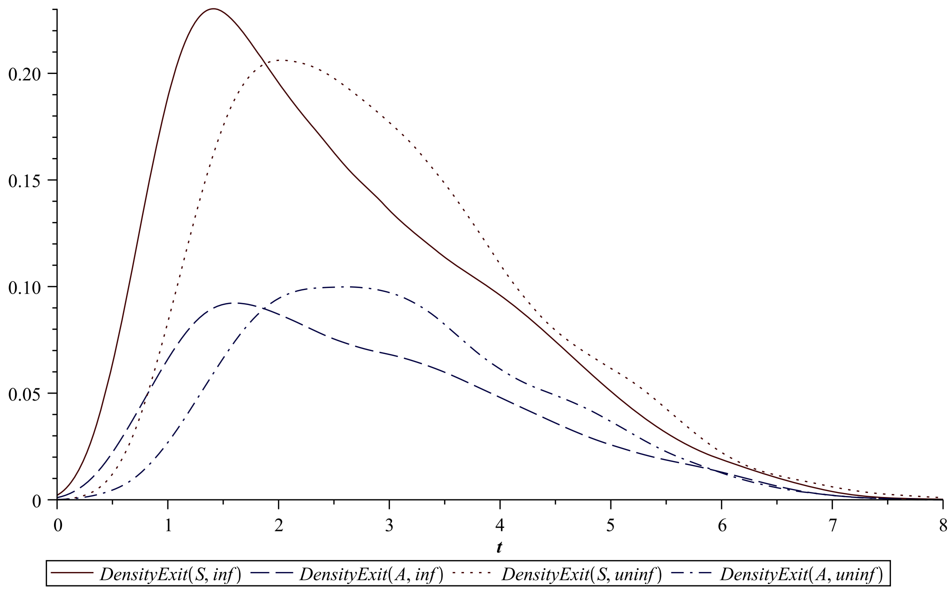
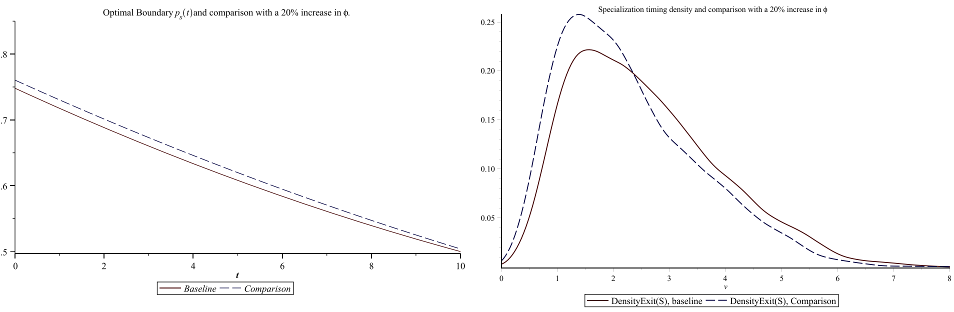



Figure 20: Specialization timing densities with informative and uninformative initial beliefs.

21 shows how the boundary and the exit density change when the precision of the signal increases by 20%. On one hand, the increased demand for experimentation (left panel) implies that agents should specialize later. On the other hand, a higher signal-to-noise ratio implies that agents reach any given target belief faster, which is the dominant effect and explains why the density of exit times puts more weight on early realizations. Notice that the two effects go in the same direction with respect to the probability of correct eventual specialization: not only the optimal boundary shifts upwards, but also agents exit faster, hence are more likely to specialize upon reaching a high confidence threshold.

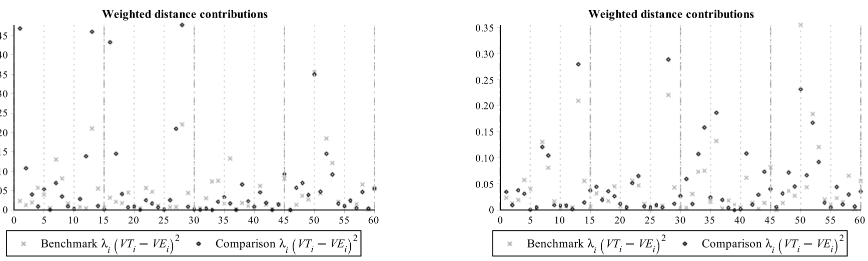



boundary $p_s(t)$ Specialization timing density ($S$ exit)

Figure 21: Change in the precision of learning, $\phi$

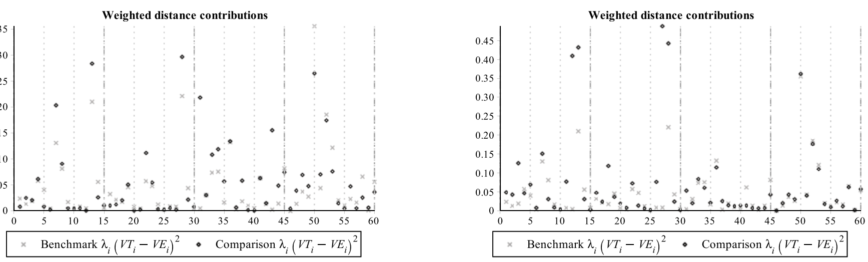
## Appendix D. Identification

Figures 22 - 26 show the weighted contribution of each moment to the distance criterion, for each of the experimental conditions discussed in Section 4.6.



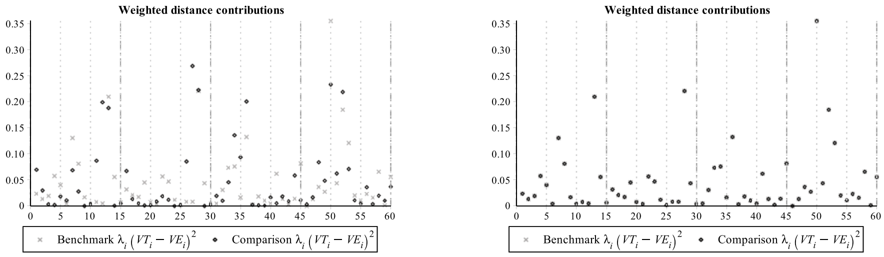

Low $\rho$ High $\rho$

Figure 22: Contribution to distance criteria: baseline (grey) and constraint (black)



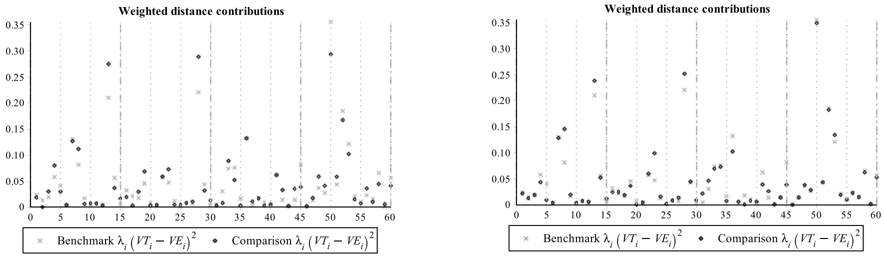

Low $\rho_0$ High $\rho_0$

Figure 23: Contribution to distance criteria: baseline (grey) and constraint (black)



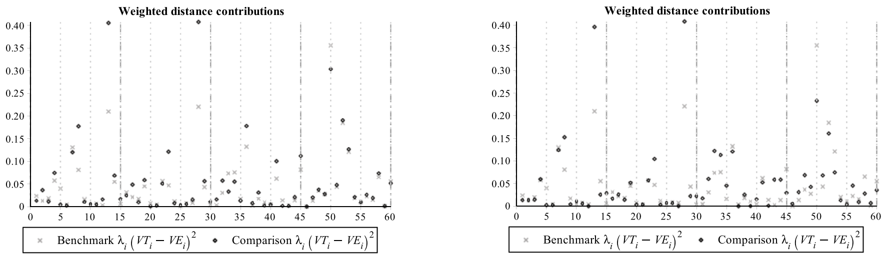

Low $\beta$ High $\beta$

Figure 24: Contribution to distance criteria: baseline (grey) and constraint (black)




Low $P$ High $P$

Figure 25: Contribution to distance criteria: baseline (grey) and constraint (black)




Low $C_\lambda$ High $C_\lambda$

Figure 26: Contribution to distance criteria: baseline (grey) and constraint (black)

# Automatic Promotion in Primary School: Immediate and Cumulated Effects on Schooling*

Margaret Leighton†, Priscila Souza‡and Stéphane Straub§

May 18, 2015

**Abstract**

This paper exploits extensive variation in grade repetition practices in Brazil to study the effect of automatic promotion cycles on grade attainment and academic persistence of primary school children. The dynamic policy environment allows us to explore the impact of automatic promotion at different times during schooling, both in the short term and as children exposed to the policy progress through primary school. We demonstrate that automatic promotion increases grade attainment: one year of exposure to the policy is associated with 3 students out of 100 studying one grade level above where they would be absent the policy. This effect persists over time, and cumulates with further exposure to the policy. It is achieved through immediate decreases in repetition rates, accompanied by compensating increases in promotion rates, particularly in early primary. While repetition rates display some degree of reversion in subsequent years, these effects are small with respect to the initial decrease. Episodes of automatic promotion have no global effect on either dropout rates or on total enrollment. Despite this, considerable sorting takes place between schools in response to the policy, with automatic promotion attracting students in younger grades and decreasing enrollment at older ones.

*We thank Sergio Firpo and Mattia Girotti for valuable insights and discussions throughout this project, and Julia Bird for contributions to the dataset. We gratefully acknowldege the *Instituto Nacional de Estudos e Pesquisas Educacionais* for sharing restricted-access data necessary to complete our panel.

†Toulouse School of Economics, m.leighton@gmail.com.

‡Climate Policy Initiative/Pontifical Catholic University of Rio de Janeiro

§Toulouse School of Economics

# 1 Introduction

While the retention of under-performing students has been a popular education policy for decades, the empirical and theoretical evidence on the practice is mixed. Two main arguments are put forward in favour of retention: on the one hand, automatically promoting students can discourage hard work by alleviating incentives, while on the other hand, struggling students may be unprepared to learn more advanced course material if they have not yet grasped the basics. These arguments suggest that grade retention, by both motivating students and assorting them to the proper level of coursework, should increase achievement.

Nevertheless, many countries have moved away from the practice in recent decades due to concerns that retention is a powerfully demotivating experience. Retained students, interpreting being held back as a negative signal about their own innate abilities, may decide that putting in effort at school is not worth the trouble. These students, who also lose their peer group as their non-retained friends are promoted, may develop problem behaviours in the classroom. Those who argue against grade retention also point to the modest learning gains of retained students, raising questions about the linearity of children's skills acquisition.

Brazil's large-scale experimentation with grade repetition provides a unique opportunity to study the impact of such policies on elementary school children. To combat the low grade attainment resulting from Brazilian students repeatedly repeating early grades of primary school, the federal government introduced a policy of automatic promotion cycles: groups of grades during which students would advance automatically to the next level without the threat of repeating. The cycles policy, while clear in intent, was inconsistent in its adoption and application. In some states, all schools adopted the policy under similar guidelines; in others, discretion over adoption was heavily decentralized. Many schools who adopted the policy later abandoned it, while other schools which had not taken it up at first did so later. While first grade was generally included inside a cycle (i.e., promotion was made automatic at the end of first grade), the length and number of cycles varied after that.

This paper exploits this variation to address the question of how grade repetition affects primary-aged children's progress through school. The dynamic movement of municipalities in and out of the policy provides within-municipality variation in exposure to the policy. Mobilizing data on the universe of Brazilian primary schools, we are then able to estimate the impact of lagged values of the policy, while simultaneously purging contemporary estimates of confounding past exposure.

We first demonstrate that the policy did indeed decrease repetition, and document the compensating increases in passing rates. We then show that the policy had no effect on overall dropout rates or enrollment. Notwithstanding this absence of attrition, we find significant and lasting increases in grade attainment for cohorts exposed to the policy. One year of exposure increases the average grade level of the cohort by approximately 2.5% - in other words, two to three students out of a hundred are studying at a grade level above what they would be otherwise.

Despite the prevalence of grade repetition as an education policy, high quality empirical research on the issue has only emerged recently. Early research into the effect of grade repetition was based on the wide-spread practice of teacher-initiated retention, particularly in the United States, and is therefore plagued with selection issues. Using data from a recent representative panel of US children, several papers assess the impact of retention in kindergarten. Hong and Raudenbush (2005) use propensity score matching on covariates to conclude that retention is harmful to retained students, and affects neither promoted students nor overall achievement. Using the same dataset, Dong (2010) uses a double hurdle model to explicitly control for selection into a kindergarten which practices retention, in addition to selection into retention itself. She finds a positive but decreasing effect of retention in kindergarten on math and reading test scores, measured in grades 1 and 3. Alet (2011) finds a similar trend in French data; using a longer panel, she finds that the effect on test scores eventually becomes negative. Estimating test scores and repetition probability simultaneously, repetition in grades 1 or 2 has a positive effect on test scores in grade 3; however, by grade 6 the positive effect of these early retentions has reversed.

In the late 1990s, Chicago moved away from a system of automatic promotion by introducing high stakes exams in grades 3 and 6, accompanied by remedial summer school for students with poor exam performance. Using regression discontinuity around the retention cut off, Jacob and Lefgren (2004) find that both aspects of the program, separately and in combination, have positive effects on subsequent test scores for students facing retention in grade 3, but no significant effects for students facing retention in grade 6. Employing the same data but a different technique, Roderick and Nagaoka (2005) reach a more pessimistic conclusion. The authors use a learning model to estimate deviations from a common trend, as well as a two-stage probit, to conclude that retention was neutral for students in grade 3, but actually harmful to students retained in grade 6. Evaluating a similar program in Florida, Greene and Winters (2007) find that retention in the 3rd grade (plus a suite of supplementary supports to retained students) increases grades two years later. In contrast to other studies, the authors find the positive effect of retention (plus supports) to be increasing: very small after the first year, substantial and significant after the second.

Which is more discouraging for a struggling student, repeating a grade or being promoted to more advanced coursework without proper preparation? Using data from grade 8 students in Chicago after the implementation of the high-stakes testing program, Allensworth (2005) finds that dropout increased among retained students, while for larger group of non-retained students, dropout decreased. André (2009) and Glick and Sahn (2010) conclude that grade repetition among primary school students in Senegal increases the risk of dropping out. Manacorda (2012) presents similar findings from Uruguay. Using a discontinuity in repetition rules for seventh to ninth-grade students, he finds that repetition increases dropout rates and reduces school attainment.

In the work most closely related to our own, Koppensteiner (2014) considers the effect of a school having a repetition policy on the test scores of 4th grade students in state-run schools in the Brazilian state Minas Gerais. The author uses schools who were late adopters of the state automatic promotion policy as treatment schools, and compares them with earlier adopters. He finds that the move to automatic grade promotion (in 2nd and 4th year) in Minas Gerais led to a 7% of a standard deviation decrease in test scores measured in 4th year.

We contribute to the literature on grade repetition in three ways. First, in contrast to much of this literature, we provide evidence on the effect of *removing* the possibility of repetition, in a context where positive repetition policies are the norm. There is good reason to think that introducing repetition in a system where automatic promotion is expected will not have a symmetric effect to removing the possibility of repeating when repeating is common. Recent studies of programs introducing repetition in the United States (see Jacob and Lefgren (2004), Roderick and Nagaoka (2005), Greene and Winters (2007)) are therefore poor guides for the many countries currently struggling with high rates of repetition. This paper addresses that problem directly, and evaluates the efficacy of introducing automatic promotion in such a context.

Second, we estimate the impact of the automatic promotion policy on children's outcomes in a developing country context. We demonstrate the success of Brazil's cycles policy with respect to grade progression, and show that these improvements are sustained.

Finally, we document the impact of the automatic promotion policy not only in the current year, but also over time. Taking twelve-year-old children as an example, we are able to estimate the impact of the policy at age twelve, but also the impact at age twelve of having been exposed to the policy at each age from seven to eleven. To our knowledge, we are the first paper to present results on the age-specific effects of automatic promotion at this level of saturation.

The remainder of paper proceeds as follows. In Section 2 we provide background information on the education context of Brazilian primary schools, and describe the cycles policy in detail. Section 3 describes the data. Section 4 presents our empirical strategy. Sections 5 present our main results and robustness exercises, respectively. Section 6 discusses the findings, while Section 7 concludes. Additional details on the data and the cycles policy, as well as robustness exercises around the primary empirical specifications, are in the Appendix.

# 2 Background

## 2.1 Education in Brazil

High-quality education is crucial both for individual success and for a country's economic and social development. Deficiencies in both the quantity and the quality of Brazilian public education have long been seen as an one of the major obstacles to growth and social inclusion in the country. Among the most important historical problems of the Brazilian educational system are the age-grade gap and the number of school-age children out of school, both of which are particularly predominant in poor and rural areas.

Although historically an important issue in Brazil, the number of school-age children out of school had declined rapidly over the past two decades. While in 1992 13.4% of children between the ages of seven and fourteen were out of school, this number fell to 3.5% in 2001, 2.3% in 2007 and 1.5% in 2013 (PNAD/IBGE).[1] Cardoso and Verner (2006) confirm this trend. Reporting on a survey of twelve- to eighteen-year-olds living in favelas of Fortaleza, the authors find that, even among this high-risk population, almost all twelve-year-olds attend school. Attendance rates start to fall at age thirteen for boys (down to 80 percent), while they remain high among girls until age seventeen. Nationwide a smaller, but still impressive, decline in the out-of-school rate can be seen among older students compared to primary-school-aged children. For those between the ages of fifteen and seventeen, the rate fell from 40.3% in 1992 to 15.8% in 2013 (PNAD/IBGE).

The proportion of students who are too old for the classes they are attending has also decreased. As can be seen in Table 1, in 1982, 72% of first graders were too old for that grade. This problem spread over all grades, so that by seventh grade 80% of students were not at the appropriate grade for their ages. The issue has been drastically reduced, but it is still important. In 2010, 15% of first graders and 29% of seventh graders were above the target age for those grades. A number of policies have contributed to this improvement in educational outcomes, including the *Bolsa* conditional cash transfer programs, introduced in the early 2000s. Glewwe and Kassouf (2012) use data from the Brazilian school census to study the impacts of *Bolsa Escola* (later *Bolsa Familia)* on enrollment, dropout and promotion rates. They find positive effects of the program on all three indicators.

While enrollment rates and age-grade misalignment have both improved substantially in recent years, they remain significant obstacles to education quality in Brazil. Both of these problems are linked to the high levels of repetition predominant across the country. Following the 2012 PISA evaluation, the OECD (2013) noted that repetition rates in Brazil remain among the highest in surveyed countries. While there was a decline in grade repetition during primary school between the 2003 and 2012 evaluations, repetition rates in secondary school increased over the same period.

Grade repetition has been the object of study in Brazil for many years. Using microdata

[1] PNAD is the National Household Sample Survey (Pesquisa Nacional por Amostra de Domicílios) from the Brazilian Institute of Geography and Statistics (IBGE-Instituto Brasileiro de Geografia e Estatística). The survey collects annual data on the characteristics of the population with a sample size of over 150,000 households.

collected by the World Bank from 1981-1985 in Northeast Brazil, Gomes-Neto and Hanushek (1994) are able to follow individual students over several years. Although their data are quite limited, the authors find that academic performance and the existence of higher grades at school are both strong determinants of grade repetition: in other words, repetition was based on commonly accepted criteria. They show that students do learn when they repeat grades, but suggest that repetition is a costly way of achieving these small gains. More recently, Koppensteiner (2014) finds that the shift towards automatic promotion cycles in Minas Gerais was accompanied by a decrease in 4th grade test scores, suggesting that repetition is indeed promoting student achievemenmt.

## 2.2 The Cycles Policy

Until the early nineties, Brazilian schools followed the practice of allowing the repetition of students at every grade level. Students could not only repeat in every grade, but could also be retained several years in a row at the same level. Starting in 1997, a number of Brazilian municipalities and states adopted a system of 'learning cycles': groupings of school grades during which promotion is automatic. For example, if the first 8 years of primary school are grouped into 2 cycles of 4 years, then students will pass 1st, 2nd and 3rd grade automatically (subject to a minimum attendance rate), but may repeat the 4th grade. They will also pass the 5th, 6th and 7th grade, but may be retained in the 8th grade. This policy is called 'Continued Progression,'[2] referred to here as the cycles policy.

This policy of learning cycles was nationally recognized in the Law of Guidelines and Foundations for Education[3] enacted by the Federal Government in 1996. This law granted additional autonomy to Municipalities and States to organize the schooling system. Although municipal schools in São Paulo had experimented with cycles as early as 1992, the first large-scale adoption of the policy was by the state of São Paulo in 1997. The Federal District and several of the 26 Brazilian States followed at various times, including Amazonas, Ceará, Espírito Santo, Mato Grosso, Minas Gerais, Paraná, Pernambuco, Rio de Janeiro and Rondônia.

[2] Progressão continuada.

[3] Lei de Diretrizes e Bases da Educação (LDB).

Adoption of the policy took a number of different forms. While in some states and municipalities the adoption of cycles was mandatory, in others the system of cycles was only recommended (but not mandatory). Therefore, in some states and municipalities schools could choose whether to adopt the automatic promotion system. While many states did not adopt the policy at all, others adopted it for some years and then retracted it. We will return to the policy and present some descriptive statistics on its adoption in Section 3.3.

# 3 Data

## 3.1 Construction of the Panel

The data which provides the starting point of this paper is the *Censo Escolar*, an annual census of schools in Brazil below the tertiary level. The survey, carried out in May, covers both private and public schools, and has been running continuously since 1995. The data are publicly available from the *Instituto Nacional de Estudos e Pesquisas Educacionais* (INEP, the national education research institute). From 1995-2006 the *Censo Escolar* measured school-level variables, whereas from 2007 on data is presented at the student level, with associated school and teacher files.[4] The *Censo Escolar* survey varies from year to year; however the general topics remain fairly consistent over time. The survey sections include basic information, physical and instructional features of the school, teachers and staff, numbers of classes and students, and student flows from the previous year (retained, passed, dropped-out, and in some years transferred).

For the purposes of our study, we merge the data from the *Censo Escolar* into two panels, one at the school level, and one at the municipality level. Each panel begins with the first data on the cycles policy, in 1999, and runs until 2006. The two panels are presented individually below.

[4] School identifiers, as well as student-level identifiers, are encrypted in the publicly-available data. This encryption prevents the identification of individual schools, but also the linkage of schools across years.

### 3.1.1 School-level Panel

The *Censo Escolar* surveyed 248,257 schools in 2004. We restrict our attention to students in grades[5] 1-8, and to the schools in which they are enrolled. This leaves 166,505 schools. Of these, 116,209 are under municipal jurisdiction, 31,178 are under state jurisdiction, 19,078 are private schools, and the remaining 40 are federally-run. Table 2 gives the mean number of schools per municipality across the panel, both overall and by administrative jurisdiction.

Not all primary schools offer both junior and senior primary classes. Table 3 lists the mean number of students enrolled in each grade, conditional enrollment being positive. As can be seen in the last column of the table, there are approximately three times as many schools offering junior primary grades as senior primary grades. Senior primary schools therefore enroll more students.

The school level panel is highly unbalanced. A total of 216,429 primary schools appear in our 8-year panel; 14,227 are only active in a single year, while less than half, 129,942 schools, are present throughout. Because our school-level regressions contain fixed effects, schools which are active only in a single year will drop out of our panel; we do not make further restrictions and retain the remaining schools for analysis.

### 3.1.2 Municipality-level Panel

Secondly, we aggregate the school-level variables at the municipal level. We aggregate the data by summing the observations across schools. This is done in such a way that it is as if the municipality had only one school, with all the students and resources pooled together. Data are then merged to create an 8-year panel with municipality-years as the unit of observation. The 1999-2006 panel is highly balanced: compared to 2006, there are 4 fewer municipalities in 2001-2004, and 57 fewer in 1999-2000. We exclude these 57 municipalities from our study.

We augment the municipality panel with census data on municipal population and gross domestic product from the *Instituto de Pesquisa Economica Applicada* (IPEA). Additional data regarding

[5]We use the word 'grade' as an analog to the Brazilian term *serie*, with corresponding levels 1-8. During our panel, the new 9-year *ano* grade-level system began to be rolled out. The extra year (*ano 1*) essentially advanced primary school enrollment by one year. Throughout this study, we abstract from any differences in the two systems beyond their duration, and convert the *ano* grades (1-9) to their *serie* equivalent, where *ano* 2 = *serie* 1. The grade *ano* 1 is excluded from the analysis. We will refer to these school years collectively as 'primary school'.

ages of children surveyed in the year-2000 census were acquired from the *Sistema IBGE de Recuperação Automática* (SIDRA).[6]

## 3.2 Outcome Variables

Our primary outcome measures are enrollment, grade attainment of those who are enrolled, and flows of students from one year to the next. The definition of these variables, and some summary statistics, are presented below.

Table 4 gives a flavour of the data, listing municipality-level average enrollment and grade attainment of a single birth cohort over six years. Data is restricted to students six to twelve years old who are enrolled in grades that are no more than two years ahead of the age-appropriate level.[7]

### 3.2.1 Enrollment

We follow Glewwe and Kassouf (2012) by using the natural logarithm of student numbers as our primary measure of enrollment. Table 5 presents summary statistics of school-level age-specific enrollments, in levels and in natural logs. We maintain two other measures of enrollment for comparison: enrollment in levels, and enrollment as a share of relevant age category from the 2000 census (see Appendix A.1.2).

### 3.2.2 Student Flows: Passing, Repetition and Drop-out

Each wave of the *Censo Escolar* collects data on the student flows from the previous year. Specifically, schools are asked to report how many students from each grade repeated or were promoted at the end of the year, and how many dropped out before the end of the year.[8] In order to convert these student counts into rates, we divide the counts by the number of students enrolled in each

[6]Data come from the section on education, accessed through: www.sidra.ibge.gov.br/bda/popul.

[7]Student ages are reported by grade. Data are available on children 'younger than six' enrolled in grade 1, but we omit these (very few) individuals because we cannot precisely determine their age. Prior to 2003, student ages were not reported for ages younger than two years below the age-appropriate grade level (e.g. number of six-year-olds is reported in grades 1 and 2, but not in higher grades). From 2003-2006 we maintain this truncation for consistency.

[8]In some waves there are additional categories: conditional or unconditional pass, transferred out of the school, joined the school part-way through the year, etc. We focus on these three because they are both the most interesting to us, and those which are most consistently measured across years.

grade, reported in the *Censo Escolar* of the previous year. We then adjust for the lag in data collection, and attribute the rates to the previous year: repetition rates in 2003 are therefore calculated based the outcomes of the 2003 school year.

Table 6 gives school-average repetition, pass and dropout rates, both overall and by policy status. Note that the difference in repetition rates between school which have adopted the cycles policy and those that have not varies considerably across grade levels. The equivalent table at the municipal level (see Table 7) presents a very similar picture: if anything, the differences between repetition rates in adopting and non-adopting municipalities are larger than the inter-school differences. The variation over time for a subset of these statistics is presented in Appendix A.2 (see Table 36).

#### 3.2.3 Grade Attainment

Measuring grade attainment presents one major limitation: grade information is available only for those students enrolled in school. Since students who are not enrolled are likely to be lower-achieving than those who are, our estimates of grade attainment should be thought of as an upper bound.[9] To calculate the mean grade attainment for a given birth cohort, we simply multiply the number of students of that year in each grade by the grade level, and divide by the total number of students born in that year. Equation (1) formalizes this approach, where $g$ is a grade level, and $n_g$ the number of students enrolled at that level.

$$E^a_{ijt} = \frac{\sum^g_{1...8}(n_g * g)}{\sum^g_{1...8} n_g} \tag{1}$$

Table 8 gives the municipality average grade of students, by age, for each year of the sample. Note that the target level for seven-year-olds is grade 1, and that average grade would increase by 1 each year if all students were promoted. If all student advanced on schedule, the mean grade for twelve-year-olds would be 6. While twelve-year-olds remain, in 2006, more than one year behind, there is steady improvement in this measure over the 8 years of the panel: the mean grade in 2006 is higher than in 1999 at all ages above seven.

[9]The fact that most children of primary school age are enrolled in school at this time attenuates this issue; however, especially for the younger children in our study (who may not have yet enrolled), this is important to bear in mind.

While we will study the effects of grade repetition on mean grade level both at the municipality and at the school level, it is important to keep in mind the significant difference between the two: while population at the municipality level changes only slowly, populations of students within schools can change much more easily - and can do so in response to the adoption of policies at the school level.

### 3.3 Policy Variable

Data on adoption of automatic promotion cycles are available in the *Censo Escolar* in two forms. In 1999, and again from 2003 to 2006, the data include individual schools' reported "total number of cycles and duration of each cycle".[10] From 2009 onward, schools are simply asked whether or not elementary school is organized in cycles.[11] While the questionnaires from 2000-2002 also contain the cycles module, the data are absent from the publicly available data files. Cycles data from these years were provided to us on request; however, these supplementary data are only yes/no. For the 1999-2006 panel, therefore, we have a consistent binary measure of cycles adoption, but no details on duration or timing of these cycles: we therefore restrict ourselves to a binary adoption variable.[12]

Summarizing the adoption of Cycles policies at the municipal level requires an aggregation which is less natural than that done for student outcomes. We first calculate the weighted share of schools within a given municipality which report using cycles: this corresponds to the probability that a randomly selected student is enrolled in a cycles-using school.[13] We then simplify the treatment of the cycles policy into a binary variable. Our intent here is compare municipalities where a majority of schools are using cycles to those where this is not the case. The distribution of the share of schools using cycles within a given municipality is highly bimodal (see Figures 1 and 2 in Appendix

[10]From the 1999 *Censo Escolar* questionnaire: *Número Total de Ciclos e Duração de cada Ciclo.*

[11]From the 2009 *Censo Escolar* questionnaire: *Ensino Fundamental organizado em ciclos.*

[12]Over the course of our panel, there are approximately 30,000 school-year observations which return a missing policy. A few examples from the data lead us to believe these are either clerical errors or misunderstandings, and are meant to indicate absence of cycles. One such example is the state of Minas Gerais where, between 2002 and 2003, the number of schools responding to the question falls by 61%, while the share of schools adopting cycles rises from 41% to 99%. This situation persists in 2004, before reverting to pattern much more similar to that observed in 2002. Coding these missing values as zeros also makes the school-level policy consistent with the municipal aggregation: weighted share of schools in the municipality reporting using cycles.

[13]Weights are calculated based on each school's enrollment of students in grades 1-8.

B), motivating our use of a binary specification. We therefore define Cycles use in a municipality as equal to one if the share of weighted share of schools reporting use of a cycles policy is greater than 75%.[14]

The annual means of our Cycles variable are shown in Table 9, both overall and for schools operated by each of the two primary public jurisdictions. The fairly stable mean prevalence of the cycles policy masks substantial volatility in the policy's adoption. As can be seen in Table 10, about 5% of municipalities move in or out of the policy every year: these municipalities go from either almost complete adoption to abandonment of the policy - or vice versa. A more detailed summary of the prevalence of the Cycle policy across the five regions of Brazil, and how this changes over time, is given in Appendix B.2.

# 4 Estimation Strategy

## 4.1 Estimating Equation

Our interest in this paper lies in identifying the effect of the cycles policy on student progress through school. We measure progress with five different variables: three student flows (repetition, promotion and mid-year dropout), total enrollment, and average grade by age. We will examine these outcomes at two different units of observation (the municipality and the school), with separate regressions on students at different grade levels (for the flow variables) or ages (for enrollment and average grade level). A first approach to this problem would be to estimate equations of the form:

$$Y_{it}^{\tau} = \alpha_0 + \alpha_1 D_{it} + \alpha_2' \mathbf{X}_{it} + \theta_i + \epsilon_{it}, \text{ for } \tau \in \{g = 1..8, a = 7..12\} \tag{2}$$

where $Y_{it}^{\tau}$ is the outcome of interest in unit $i$ at time $t$ for students $\tau$ (with $\tau$ representing either the grade-level or age group, as appropriate); $D_{it}$ is a dummy for the policy; $\mathbf{X}_{it}$ is a vector of time-varying characteristics at the unit level; $\theta_i$ is the unit fixed effect and $\epsilon_{it}$ is the unobserved error term. Note that we consider outcomes separately by age or grade group, so the outcomes

[14] We also compute a series of alternate thresholds, retained for robustness exercises in Section C.3. The alternate measures are summarized in Tables 40 - 42, Appendix B.1.

$Y_{it}^{\tau}$ will be, for instance, the repetition rate of 3rd grade students, or the average grade of enrolled ten-year-olds. In each case we are interested in estimating the coefficient $\alpha_1$: the effect of exposure to the policy in the current year on the outcome under consideration.

While there is some movement in and out of the policy every year, the adoption of cycles remains highly persistent over time (see Table 10). It is unlikely that the history of the policy in a given school or municipality will be too important for young students enrolling in school for the first time.[15] For older students, however, who have been in school for several years, past exposure to the policy could affect their progress through school today. Given the persistence of the policy, if we fail to control for this past exposure we run the risk of overestimating the effect of the policy in the current year: students exposed today were likely also exposed last year.

As a second approach, we therefore estimate an analog of Equation 2 including lagged values of the policy. For outcomes by grade, we select the number of lags in order to coincide with the years students would have been in school, had they been progressing at the target rate. For instance, for the outcomes of grade two students, we include the current and lagged value of the policy, while for grade three students an additional lag is added.[16] When considering outcomes by age, where early enrollment is a result of interest, we include lagged values of the policy back to age six. In practice, we include one lag for seven-year-olds, two for eight-year-old, and so on. Since we are constrained in the number of years for which we have data on the cycles policy, we include a maximum of five lagged values of the policy at the upper grades and ages.[17]

This second approach can be summarized in the following equation, where $\mathbf{D}_i$ is a vector containing both current and past policy dummies:

$$Y_{it}^{\tau} = \beta_0 + \beta_1' \mathbf{D}_i + \beta_2' \mathbf{X}_{it} + \theta_i + \epsilon_{it}. \tag{3}$$

[15]There could of course be some impacts: the school could be better adjusted to the policy if they have had it for several years, or students may enroll earlier or later as a result of the policy.

[16]This approach does not, unfortunately, include all years that all students in the group in question have been at school: if a student in second grade repeated last year, he is currently in his third year of school and therefore one year of his policy history is not controlled for. On the other hand, including additional lags decreases the number of years on which our data can be estimated. We adopt this approach as a compromise between the two.

[17]We limit our age-based analysis to students aged seven to twelve, with the constraint binding only for twelve-year-olds. Since we look at grade-specific outcomes up to grade eight, this limit is hit for most of upper primary school.

More explicitly, taking as example an outcome which we examine by age group ($\tau \in (7..12)$), we estimate the series of equations given in (4) below.

$$
\begin{aligned}
Y_{it}^{7} &= \beta_0 + \beta_{10}^{7} D_{it} + \beta_{11}^{7} D_{it-1} + \beta_2' \mathbf{X}_{it} + \theta_i + \epsilon_{it} \\
Y_{it}^{8} &= \beta_0 + \beta_{10}^{8} D_{it} + \beta_{11}^{8} D_{it-1} + \beta_{12}^{8} D_{it-2} + \beta_2' \mathbf{X}_{it} + \theta_i + \epsilon_{it} \\
Y_{it}^{9} &= \beta_0 + \beta_{10}^{9} D_{it} + \beta_{11}^{9} D_{it-1} + \beta_{12}^{9} D_{it-2} + \beta_{13}^{9} D_{it-3} + \beta_2' \mathbf{X}_{it} + \theta_i + \epsilon_{it} \\
Y_{it}^{10} &= \beta_0 + \beta_{10}^{10} D_{it} + \beta_{11}^{10} D_{it-1} + \beta_{12}^{10} D_{it-2} + \beta_{13}^{10} D_{it-3} + \beta_{14}^{10} D_{it-4} + \beta_2' \mathbf{X}_{it} + \theta_i + \epsilon_{it} \\
Y_{it}^{11} &= \beta_0 + \beta_{10}^{11} D_{it} + \beta_{11}^{11} D_{it-1} + \beta_{12}^{11} D_{it-2} + \beta_{13}^{11} D_{it-3} + \beta_{14}^{11} D_{it-4} + \beta_{15}^{11} D_{it-5} + \beta_2' \mathbf{X}_{it} + \theta_i + \epsilon_{it} \\
Y_{it}^{12} &= \beta_0 + \beta_{1}^{12} D_{it} + \beta_{11}^{12} D_{it-1} + \beta_{12}^{12} D_{it-2} + \beta_{13}^{12} D_{it-3} + \beta_{14}^{12} D_{it-4} + \beta_{15}^{12} D_{it-5} + \beta_2' \mathbf{X}_{it} + \theta_i + \epsilon_{it}
\end{aligned} \tag{4}
$$

The coefficients of interest, the $\hat{\beta}_{10}^{\tau} - \hat{\beta}_{15}^{\tau}$, provide estimates of the effect of the policy in each year, controlling for past exposure to the policy.

## 4.2 School-level Analysis

Our finest unit of observation, both for our outcome variables and the policy itself, is at the level of the school. The identifying assumption for school-level regressions is that there are no unobserved, time-varying factors at the school level that are correlated both with policy adoption and with the outcomes we measure. Given that our student outcomes are measured only on those students enrolled, the possibility of students choosing their school based on the policy is a real concern.

Anecdotal evidence suggests that parents were displeased with the policy, and were concerned that students would under-perform if they did not face the threat of repetition. To the extent that this is true, more motivated parents may shift their children to schools which have not adopted the policy – possibly even sending their children to private schools to achieve this. Even in the United States, where repetition may not be considered an amenity by parents, Dong (2010) shows that there is significant positive selection into schools with repetition in kindergarten.

Despite this concern, we include school-level regressions in our results. This allows us, first of all, to discuss our results in relation to previous studies which have been done at the school level, such as Koppensteiner (2014) and Glewwe and Kassouf (2012). Comparing results at the municipal and school levels also enables us to discuss what biases would emerge if we looked exclusively at the school level, and provides some evidence on the magnitude and timing of student flows in response to the policy.[18]

## 4.3 Municipality-level Analysis

Aggregating our data at the municipality level allows us to address the student selection issue. Moving children to schools in a different municipality would be extremely costly. While families living on a municipal boundary may do so in response to the policy, it seems unlikely that this practice would be very widespread.

Our municipal policy variable does, however, have limitations. First, it is a noisy aggregation. There may be some schools in the municipality who do not adopt cycles, even if the municipality is coded as an adopter. These non-adopters are unlikely to be randomly drawn. Private schools, for instance, rarely implement automatic promotion cycles. Similarly, public schools who defy the municipal norm are likely to have special characteristics. Second, adopting a policy of automatic promotion cycles does not mean promoting students automatically in each grade. As described in Section 2.2, some grades were commonly included in cycles, while promotion from others were rarely - if ever - accorded automatically.

Both of these issues introduce measurement error into our policy variable. We expect the first to dilute any effect of the policy, as some schools are opting out, and bias our coefficient estimates towards zero. The second prevents us from interpreting treatment at the school or municipal level as a treatment at any specific grade level - with the possible exception of grade 1. As we will see when we look at the effect of the policy on repetition rates, the implementation of cycles varied

[18]One additional caution is necessary when interpreting school-level results. We will consider outcomes of two different groups of students: age-specific outcomes, and grade-specific outcomes. Because not all schools teach all grades, the interpretation of the coefficients on the vector of lagged policy values, the $\beta_1$s, must be carefully considered. While we know the history of policy adoption for all schools, students who enroll at a senior primary school for the first time in grade 5 have not been effected by the policy history at that school.This is not the case for junior primary schools.

considerably by grade level.

The identifying assumption at the municipality level is that changes in policy adoption are uncorrelated with any time-varying unobserved municipal characteristics which are, in turn, correlated with our outcome variables.

## 5 Results

The following subsections present results from regressions described in Equations 2 and 3. All regressions include fixed effects (at either the school or municipality level), state-year interaction dummies, and a collection of time-varying controls. For schools, these controls include: a dummy for location (rural or urban),[19] a dummy for jurisdiction (state, municipal or private), total number of teachers, and number of teachers at the primary level. For regressions using students aggregated by grade, the number and education of teachers teaching either grades 1-4 or 5-8 are also included.[20] The additional controls included in municipal regressions are similar: the number of schools in the municipality, the number of schools by location and jurisdiction, population and municipal GDP in natural logs and in levels, and the number and education scores of teachers teaching grades 1-4 and 5-8. Summary statistics for controls at the school and municipal levels can be found in Appendix A.3 (see Tables 37 and 38).

We analyise the effect of the policy on outcomes by age or by grade, depending on data availability. The flows of student from one year to the next are only available by grade: in other words, we have repetition rates, passing rate and dropout rates for grade 3 students, but not for ten-year-olds. For enrollment and grade level, however, we have data by age. While this difference does not greatly affect our interpretation of estimates $\hat{\alpha}_1$ from Equation 2, it does affect our interpretation of the dynamic effects, the $\hat{\beta}_1$s from Equation 3. Since the flows themselves affect the composition of each grade, lagged values of the policy are more delicate to interpret in grade-specific regressions than in age-specific regressions.

[19] Approximately 2% of schools experience a location status change during the panel. It is not clear whether the schools themselves moved, the surrounding area developed, or if these are clerical errors.

[20] For regressions by age category we do not use these controls, as not all schools enrolling students of a given age necessarily have classes at levels 1-4 and/or 5-8.

To understand why this is the case, consider average grade by age as an outcome. For students who are ten years old, the policy variable, and each lagged value of this variable, has a clear interpretation: it is exposure to the policy at ages ten, nine, eight etc. Now consider an outcome by grade, such as dropout rate. Because the group of students in a given grade in a given year is composed of promoted and repeated students from past years, lagged values of the policy variable are harder to interpret. For example, consider the dropout rate of grade 4 students. While the coefficient on the contemporary policy variable gives the effect of exposure to the policy on the dropout rate in grade 4, in previous years these students were not necessarily all studying at the same grade level. Some students were in grade 3 last year, and were promoted, while other students were in grade 4 last year, and either dropped out (and re-enrolled the next year), or completed the year and were required to repeat. We cannot, therefore, interpret the coefficient on the lagged policy variable as the effect of exposure to the policy in grade 3 on students when they are in grade 4. Rather, these coefficients capture the average effect of exposure to the policy last year for students who are currently in 4th grade. This prevents us from using our estimates to compare the long-term impact of exposure to the policy at different grades. It does, however, allow us to purge composition effects from our current-period policy dummy.

## 5.1 Student Flows

### 5.1.1 Repetition Rate

Tables 11 and 12 present results from a series of municipal-level regressions, following Equations 2 and 3 respectively, where the repetition rate at a given grade level is the dependant variable. Table 11 gives us the current year impact of the Cycle policy, for each grade, without controlling for previous policy exposure. With the exception of grade 8, policy exposure decreases repetition rates. We get a sense, however, of the uneven application of cycles across grades. Exposure to the policy decreases repetition rates by 5 percentage points in grade 1. The smaller, though still positive, coefficient on the policy in grade 2 suggests that fewer schools included second grade in a cycle, and that repetition was often possible in that grade. The effect of the policy is less than one-tenth as large in grade 4 as it was in grade 1. This coincides with grade 4 being the final year of junior primary school, and therefore unlikely to be included within a cycle of automatic promotion. Fifth grade, in contrast, as the first year of upper primary school, would be a natural candidate, and we see a substantial reduction in that year as well.

Table 12 presents results of similar regressions, augmented with lagged values of the policy. The contemporary effect of the policy changes little in junior grades; however, in senior grades the standard errors increase substantially. More interestingly, we do see some positive effects of lagged exposure to automatic promotion on current repetition rates. In column (2), we see that having been exposed to cycles in the previous year increases grade 2 repetition rates by half of a percentage point, controlling for current year exposure. We see similarly-sized, positive effects on lagged values of the policy, particularly at the lag corresponding approximately to having cycles in grade 1.[21]

Tables 13 and 14 replicate these regressions at the school level. The coefficient estimates are notably larger at the school level than at the municipal level in grade 1, and tend to be somewhat larger throughout junior primary school. Results are quite similar between the municipal and school-level saturated regressions (Table 13), although standard errors at the school level tend to be smaller, and the contemporary negative effect of the policy on repetition rates persists in all grades except grade 8.

#### 5.1.2 Passing Rate

Passing rates should follow the reverse pattern from what we observe with repetition rates. Tables 15 and 16 (at the municipality level) and Tables 17 and 18 (at the school level) show that this is indeed the case. Table 15, presenting results at the municipal level without lagged values of the policy, is a close mirror image of 11, with the only noticeable differences being the standard errors for estimates at grades 4 and 8. Comparing the equivalent two tables with lagged values of the policy yields a slightly different picture. While repetition rates respond to lagged values of the policy (see Table 12), passing rates appear not to. Closer inspection shows that the signs and sizes of the coefficients do in fact mirror those in Table 12, however, the standard errors are substantialy

[21]These are the coefficients on the lower diagonal: they correspond in time to grade 1 for students who never repeat.

larger in the passing rate regressions. School-level results on passing rates are similarly comparable to those on repetition rates at the junior level, with more exceptions in grades 6-8.

#### 5.1.3 Drop-out Rate

Tables 19 to 22 presents results of regressions which take dropout rates as the dependant variable. At the municipal level, when no lagged values of the policy are included, cycles have a small but consistently negative effect on dropout rates at most grade levels. Exposure to the policy in the current year decreases dropout rates by an average of about 0.3 of a percentage point: given that dropout rates average 7%-13% depending on the grade, this is a small but not insignificant change. When a history of lags is included in the regression, most of these effects disappear, leaving only the decrease in first grade.

Regressions at the school level reveal a very different pattern. Considering first Table 21, which presents results from regressions with only the contemporary policy variable, we can see that cycles are often associated with an *increase* in dropouts. This is particularly the case in grades in which cycles had the largest effect on repetition rates: grades 1, 3 and 5. We do see some decrease in drop out in grades 2 and 4; however, when we turn to the results from saturated regressions (see Table 22), those negative effects – particularly in 4th grade – lose significance. The increase in dropout rates in key cycles grades remains in regressions with lagged values of the policy, but no pronounced trend emerges regarding past exposure to the policy.

### 5.2 Enrollment

Tables 23 to 26 present series of regressions with the natural log of total enrollment at each age as the dependant variable.[22] Based on the municipality-wide results in Tables 23 and 24, it appears that the automatic promotion policy neither encouraged nor discouraged students from enrolling in school in general. When lags of the policy are included in the regression a slight negative effect on enrollment appears amongst older cohorts: cycles at age eleven decrease enrollment at ages eleven and twelve; however, these effects are quite small (see columns (5) and (6) in Table 24).

[22] These regressions are replicated using levels rather than natural logs in Section C.1: see Tables 45 and 44.

The school-level regressions in Tables 25 and 26 reveal that the stability of enrollment at the municipal level hides considerable movement of students at the school level. While the size of the effects remain small, the trend in enrollments is quite pronounced: automatic promotion is associated with higher enrollments of younger children, and lower enrollments of older children.

### 5.3 Grade Attainment

Tables 27 to 30 present a series of regressions with the mean grade level at each age as the dependant variable, for municipalities and schools respectively. Note that this variable is calculated based only on those students enrolled in school, therefore the minimum value is achieved when all students of that age who are enrolled are in grade 1.

Considering first the municipality-level regressions, Tables 27 and 28 show that the automatic promotion policy had a significant and lasting impact on the average grade level of each birth cohort. The absence of an effect at age seven is unsurprising: students are normally enrolled in primary school for the first time at this age, and therefore have not yet faced the possibility of repetition.[23] From age eight onwards, both contemporary automatic promotion and past exposure to automatic promotion increase average grade level. By looking at the coefficients along the diagonal, we can see that the positive effect of past policy exposure is highly persistent over time. The coefficients are also quite similar in magnitude, regardless the age of exposure, ranging from approximately 0.01 to 0.04, with a mode around 0.03. In other words, exposure to the policy during one year increases the average grade attainment of the enrolled cohort permanently, with 3 children out of 100 at a grade level higher than they would be without the policy. Coefficients from school-level regressions, reported in Tables 29 and 30 are in general slightly smaller, and in some cases even negative.

[23] There are some exceptions, as students do occasionally enroll in grade 1 at age six or lower.

# 6 Discussion

## 6.1 Schools vs. Municipalities

To interpret the regression coefficients described in the previous section, we must first address how the policy affected the composition of the student groups under consideration. Changes in enrollment would imply that the characteristics of the student body are also changing: average grade, and progression of a grade-group from one year to the next, will then be measured on a different set of students when the policy is in place. We find no evidence, however, that the cycles policy significantly increased municipal enrollment of children aged seven to ten, and only scant evidence of a decrease in enrollment at ages eleven and twelve. By and large, enrollment rates of children at these ages are already quite high, so this result is not surprising. This finding gives us confidence in interpreting results in Section 5 as true effects of the cycles policy on the progress of students, rather than compound effects which simultaneously alter the composition of the student body.

This does not appear to be the case at the school level. The adoption of a the cycles policy at the school level is associated with increases in enrollment at ages seven to ten, and decreases at age twelve (and, to a lesser extent, eleven). If lower-ability students are sorting into schools with automatic promotion, our estimates of school outcomes are likely to be downward biased; if, on the other other hand, higher ability students are opting in to policy-adopting schools, our results should be biased upwards. Comparing municipal and school results across our other outcomes gives some evidence to the former. The effect of the policy on grade attainment by age (see Tables 28 and 30) is lower at the school level than at the municipal level, suggesting that low-achieving students are disproportionately choosing cycle schools. We also find evidence at the school-level that the cycles policy increased dropout rates, particularly in those grades where cycles appear to be most frequently implemented. Nevertheless, the decrease in repetition rates in first grade is larger at the school level than at the municipality level. While this could be partly due to the improved measurement of the policy at the school level, it is worth noting that, even in the presence of the policy, repetition was still possible under some circumstances.[24] Any sorting that does take place is therefore not severe enough to overcome this basic reduction in repetition rates.

### 6.2 Main Results

Our results demonstrate, firstly, that the cycles policy was implemented as intended: repetition rates fell, particularly in first grade, and promotion rates increased. Note, however, that repetition rates did not fall to zero: even at the school level, while repetition rates in first grade fell by 8 percentage points, cycle schools still repeated 5% of first graders. Nevertheless, the policy is associated with significant reductions in repetition and increases in grade promotion.

We observe these effects in the absence of any significant changes in enrollment or dropout rates. To some extent, the lack of an effect on enrollment may reflect the improvements in schooling rates in Brazil in recent years: given that most children of elementary school age are annually enrolled in school, there is little room for improvement on that front.

The absence of a significant change in dropouts is difficult to interpret. As Allensworth (2005) demonstrates, the effects of repetition on dropout can vary dramatically between repeaters and non-repeaters, and the absence of an overall effect can dissimulate compensating changes in these two groups. If students are promoted ahead of their abilities, they may face a discouraging mismatch between their abilities and the course material.[25] If this mismatch is severe, students may feel hopeless and dropout. On the other hand, repetition itself can be discouraging, and precipitate school-leaving. Given that automatic promotion could theoretically affect promoted students in either direction, we cannot draw firm conclusions. It is noteworthy, however, that there was in fact no overall effect: observed gains in grade attainment were achieved without driving students out of the classroom.

The substantial effects of the policy on age for grade are the most striking findings of the paper. Not only are the effects non-trivial – raising on average grade attainment by 0.03 per year on average – they are sustained and cumulative. To the best of our knowledge, this is the first evidence on the

[24] For instance, if attendance drops below a certain threshold.
[25] See, for instance, Pritchett and Beatty (2012) on learning profiles.

medium-term effect of introducing automatic promotion, and it is surprisingly positive. Cohorts exposed to automatic promotion, despite facing repetition in the future, have permanently higher grade attainment through to the end of primary school.

## 6.3 How Do these Results Compare to Previous Studies?

### 6.3.1 Student Outcomes in Minas Gerais

In the work most closely related to our own, Koppensteiner (2014) studies the effect of introducing the cycles policy on test scores among students at state-run school in Minas Gerais. Studying such a restricted subset of schools has one important advantage: while the length and timing of cycles varied considerably among those schools adopting the policy across Brazil, Koppensteiner describes an implementation among state schools in Minas Gerais that left little school-level discretion. This decreases the measurement error in the policy variable substantially, since all schools were applying automatic promotion and repetition cycles along the same schedule, and allows for a more precise interpretation of the results. The shortcoming of looking only at state-run primary schools is that it prevents meaningful aggregation at the municipal level, and therefore is sensitive to students sorting themselves across schools.

Although not the focus of the paper, Koppensteiner also estimates of the effect of the policy on student flows for the two cohorts he studies. While he also finds significant, negative effects of the policy on repetition rates, in contrast to our findings these effects only appear in 2nd and 4th grades. In Appendix C.5 we replicate our student flow results using only state schools in Minas Gerais. While we find very similar estimates for 2nd and 4th grade (see Table 54), our finding of large and significant decreases in repetition rates in grade 1 (where we still see the largest effect) and grade 3 remain at odds with his results.

What could account for these differences? While the tables reported in Appendix C.5 restrict our sample to state schools in Minas Gerais, we are nevertheless estimating our model on different data sets. While Koppensteiner uses data from 2000-2006, almost identical to our own panel, he follows only two theoretical cohorts over that timeframe. In contrast, we estimate our model on all children enrolled in grades 1-8 between 1999-2006. As Koppensteiner observes, the absence of effect of the policy on repetition rates in first grade arises because the cohort in question was only treated with the policy from second grade: it should therefore be interpreted as an absence of anticipatory effects, rather than an absence of causal impact of the policy (see footnote 29, page 285). The difference in our estimates of the effect of the policy on repetition rates in grade 3 remains puzzling; however, differences in the cohorts on which the analysis was done are likely responsible.

#### 6.3.2 The *Bolsa* Programs

How does the cycles policy compare with the *Bolsa* program? Glewwe and Kassouf (2012) study the impact of *Bolsa escola / Bolsa familia* on enrollment, drop out and passing rates, both at the municipal and school levels. In their basic school-level model without lagged values of the policy, they find that the program increased enrollment by 2.8 percentage points for students in grades 1-4, and 3.2 percentage points for students in grades 5-8. While our enrollment regressions are at the age level rather than the grade level, we can nevertheless approximate a comparison by averaging the coefficient in Table 26 for children ages seven to ten (target ages for junior primary) and eleven-twelve (target ages from grades 5 and 6). Doing so, we find a 1.9 percentage points increase in enrollment for the junior ages, while the negative effect at age twelve dominates giving an average decrease of 0.4 percentage points. At the municipal level, Glewwe and Kassouf do not find any effect of the existence of the *Bolsa* program in the municipality on enrollment in younger grades, though they find a 4 percentage point increases in grades 5-8. We find no effect of cycles on municipality-level enrollment at ages six to twelve.[26]

Glewwe and Kassouf also find significant school-level decreases in dropout rates, and increases in promotion rates, due to the presence of the *Bolsa* program. They find that dropout rates decrease by 0.3% across primary grades, while promotion rates increase by 0.5% (in grades 1-4) and 0.3% (in grades 5-8). These estimates compare to our estimates of the effect of the cycles policy (again, averaged across grades) of 0.2% (grades 1-4) and 0.6% (grades 5-8) *increases* in dropout rates,

[26] We do not control for the *Bolsa* programs in our main regressions because data on the program are only available from 2001. Furthermore, by 2004 nearly every school had students on the program (see Table 39). In Appendix C.4 we replicate our main regression results while controlling for the program, and find no significant changes to our results due to the program.

and 3.3% (grades 1-4) and 1.1% (grades 5-8) increases in promotion rates.[27] The authors find no significant effect of the existence of the program in the municipality on promotion or dropout rates at the municipal level; we find little effect on dropout (besides an increase in first grade), but substantial increases in promotion rates at all grades except for grade 4.

While a thorough comparison of the two programs is beyond the scope of this study, it is interesting to note that the magnitude of the effects on the outcomes discussed above are in fact quite similar. The *Bolsa* programs increased enrollment and reduced dropout more noticeably than the cycles program, at least at the school level, while the cycles program increased promotion rates significantly more, in keeping with the goals of each program. Nevertheless, the *Bolsa* program – somewhat surprisingly – also increased promotion rates.

# 7 Conclusion

Grade repetition has historically been a popular, but poorly understood, education policy. In this paper, we exploit extensive policy variation in repetition policies in Brazil to study how the introduction of periodic automatic promotion affected grade attainment and annual grade progression of primary school children.

We find that the policy did indeed reduce repetition rates, particularly in younger grades, and brought about compensatory increases in promotion rates. Past exposure to the cycles policy increases repetition rates in subsequent years; however, these effects are modest and do not compensate for the reductions observed in earlier grades. We find no convincing evidence that the automatic promotion policy either reduced dropout, with the exception of a small reduction in first grade, or increased enrollment. Our results do suggest that considerable sorting takes place between schools in response to the policy, with automatic promotion attracting students in junior grades and driving them away at higher levels.

Our results show that adoption of the cyclical automatic promotion policy at the municipality-level increased grade attainment by one year for approximately 3 out of 100 children. This increase

[27] See Tables 21 and 17. Two coefficients below the 10% significance level are included in these averages, though they are close to zero.

persists over time, and cumulates with continued use of cycles. For example, a cohort of eleven-year-old children who have been exposed to the policy since age seven would have an average grade attainment 15.6% above an equivalent cohort with no exposure to the policy: in other words, about 16 children would be one grade ahead of their peers in municipalities without the policy.

While decreases in repetition rates, and some associated increase in average grade attainment by age, should be direct outcomes of any automatic promotion policy, the persistence of these effects over time is notable. This finding is particularly remarkable given that the cycles policy applied automatic promotion only at certain grades: those students who remain in school must nevertheless face the threat of repetition periodically throughout their schooling. The robust increase in average grade suggests that, among those students who were pushed ahead despite poor performance, at least some of them succeeded in overcoming earlier learning delays.

While these increases in grade attainment suggest that the cycles policy had a positive effect on the education of primary-school-aged children in Brazil, a natural question that arises is how this accelerated advancement affected learning itself. In follow-up work, we intend to investigate this issue by pairing the current findings with results from standardized tests taken across the country.

## References


Alet, E. (2011): "Is grade repetition a second chance?" *Mimeo*, 1–35.

Allensworth, E. (2005): "Dropout rates after high-stakes testing in elementary school: A study of the contradictory effects of Chicago's efforts to end social promotion," *Educational Evaluation and Policy Analysis*, 27, 341–364.

André, P. (2009): "Is grade repetition one of the causes of early school dropout? Evidence from Senegalese primary schools." *Working paper*, 1–34.

Cardoso, A. R. and D. Verner (2006): "School drop-out and push-out factors in Brazil: The role of early parenthood, child labor, and poverty," *Working paper.*

Dong, Y. (2010): "Kept back to get ahead? Kindergarten retention and academic performance," *European Economic Review*, 54, 219–236.

Glewwe, P. and A. L. Kassouf (2012): "The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil," *Journal of Development Economics*, 97, 505–517.

Glick, P. and D. E. Sahn (2010): "Early Academic Performance, Grade Repetition, and School Attainment in Senegal: A Panel Data Analysis," *The World Bank Economic Review*, 24, 93–120.

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

Greene, J. P. and M. A. Winters (2007): "Revisiting grade retention: An evaluation of Florida's test-based promotion policy," *Education Finance and Policy*, 2, 319–340.

Hong, G. and S. Raudenbush (2005): "Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics," *Educational Evaluation and Policy Analysis*, 27, 205–224.

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

Koppensteiner, M. F. (2014): "Automatic grade promotion and student performance: Evidence from Brazil," *Journal of Development Economics*, 107, 277–290.

Manacorda, M. (2012): "The cost of grade retention," *Review of Economics and Statistics*, 94, 596–606.

Menezes-Filho, N., R. Moita, and E. de Carvalho Andrade (2014): "Running Away from the Poor: Bolsa-Familia and Entry in School Markets," *Mimeo*, 1–43.

OECD (2013): "BRAZIL – Country Note –Results from PISA 2012," Tech. rep.

Pritchett, L. and A. Beatty (2012): "The Negative Consequences of Overambitious Curricula in Developing Countries," *Mimeo*, 1–47.

Roderick, M. and J. Nagaoka (2005): "Retention Under Chicago's High-Stakes Testing Program: Helpful, Harmful, or Harmless?" *Educational Evaluation and Policy Analysis*, 27, 309–340.

Table 1: **Percent of Students in a Grade Not Appropriate for their Age**

| Year | Grade 1 | Grade 2 | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 |
|---|---|---|---|---|---|---|---|
| 1982 | 71.9 | 76.5 | 77.2 | 76.6 | 76.6 | 80.2 | 79.8 |
| 1991 | 59.5 | 62.6 | 63.3 | 62.7 | 62.7 | 68.6 | 67.4 |
| 1996 | 40 | 44.1 | 46.4 | 46.6 | 46.6 | 53.2 | 49.2 |
| 2006 | 17.5 | 24.6 | 27.5 | 28.5 | 28.5 | 35.5 | 34.1 |
| 2007 | 18.3 | 23.7 | 27.2 | 28.2 | 28.2 | 34.4 | 32.1 |
| 2008 | 15.3 | 19.3 | 20.3 | 22.2 | 22.2 | 27.8 | 25.8 |
| 2009 | 15.4 | 21.5 | 22.5 | 23 | 23 | 29.5 | 27.5 |
| 2010 | 14.5 | 21.4 | 24 | 24.4 | 24.4 | 30.7 | 28.3 |

*Source:* PNAD, using data from the Censo Escolar

Table 2: **Mean number of primary schools per municipality**

| Year | Total | Municipal | State | Private | Federal |
|---|---|---|---|---|---|
| 1999 | 33.192 | 23.750 | 6.208 | 3.224 | 0.009 |
| 2000 | 32.838 | 23.547 | 6.027 | 3.256 | 0.008 |
| 2001 | 31.854 | 22.711 | 5.837 | 3.297 | 0.008 |
| 2002 | 30.901 | 21.828 | 5.724 | 3.342 | 0.008 |
| 2003 | 30.282 | 21.247 | 5.635 | 3.393 | 0.007 |
| 2004 | 29.814 | 20.905 | 5.517 | 3.386 | 0.007 |
| 2005 | 29.115 | 20.484 | 5.239 | 3.384 | 0.007 |
| 2006 | 28.449 | 19.915 | 5.128 | 3.399 | 0.007 |

*Source: Censo Escolar*

Table 3: **Mean number of students enrolled per school**

| Variable | Mean | Std. Dev. | N |
|---|---|---|---|
| Grade 1 | 36.045 | 49.51 | 1233454 |
| Grade 2 | 31.226 | 43.653 | 1214054 |
| Grade 3 | 30.159 | 42.665 | 1180755 |
| Grade 4 | 30.31 | 43.487 | 1114851 |
| Grade 5 | 89.472 | 90.005 | 418137 |
| Grade 6 | 81.211 | 76.407 | 389640 |
| Grade 7 | 76.745 | 74.320 | 368378 |
| Grade 8 | 73.184 | 71.726 | 347244 |

Table 4: **Municipal enrollment and grade of 1994-born cohort, by year**

| Year | grade 1 | grade 2 | grade 3 | grade 4 | grade 5 | grade 6 | grade 7 | grade 8 |
|---|---|---|---|---|---|---|---|---|
| 2000 | 82.38 | 1.78 | . | . | . | . | . | . |
| 2001 | 456.16 | 34.49 | 1.61 | . | . | . | . | . |
| 2002 | 241.65 | 345.12 | 29.45 | 1.63 | . | . | . | . |
| 2003 | 73.68 | 219.17 | 308.06 | 28.21 | 1.45 | . | . | . |
| 2004 | 32.29 | 83.73 | 196.34 | 289.20 | 26.43 | 1.40 | . | . |
| 2005 | 17.38 | 42.56 | 80.24 | 187.87 | 273.99 | 25.10 | 1.35 | . |
| 2006 | 8.59 | 23.43 | 45.30 | 83.15 | 193.66 | 247.70 | 21.53 | 1.41 |

*Source:* Censo Escolar, means by municipality

Table 5: **School-level enrollments: levels and natural log**

| Variable | Total | | Ln total | | N |
|---|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. | |
| Age 6 | 6.645 | 12.833 | 1.19 | 1.066 | 474989 |
| Age 7 | 18.628 | 28.59 | 2.108 | 1.285 | 1147125 |
| Age 8 | 22.546 | 34.779 | 2.262 | 1.318 | 1208863 |
| Age 9 | 22.651 | 34.84 | 2.258 | 1.327 | 1221614 |
| Age 10 | 22.29 | 34.585 | 2.217 | 1.346 | 1253105 |
| Age 11 | 22.394 | 32.961 | 2.213 | 1.383 | 1251465 |
| Age 12 | 23.948 | 38.664 | 2.142 | 1.47 | 1017822 |

Table 6: **School-level student flows by policy status**

| | Cycles | Grade 1 | Grade 2 | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|---|---|
| Repeat | | 0.19 | 0.15 | 0.11 | 0.09 | 0.12 | 0.10 | 0.08 | 0.07 |
| | No | 0.22 | 0.16 | 0.12 | 0.09 | 0.12 | 0.10 | 0.08 | 0.06 |
| | Yes | 0.05 | 0.13 | 0.06 | 0.09 | 0.10 | 0.09 | 0.07 | 0.08 |
| Pass | | 0.70 | 0.77 | 0.80 | 0.83 | 0.77 | 0.81 | 0.82 | 0.85 |
| | No | 0.67 | 0.76 | 0.79 | 0.83 | 0.77 | 0.80 | 0.82 | 0.86 |
| | Yes | 0.86 | 0.82 | 0.88 | 0.86 | 0.79 | 0.82 | 0.81 | 0.84 |
| Drop-out | | 0.11 | 0.08 | 0.08 | 0.08 | 0.10 | 0.09 | 0.09 | 0.08 |
| | No | 0.11 | 0.09 | 0.09 | 0.09 | 0.10 | 0.09 | 0.09 | 0.08 |
| | Yes | 0.07 | 0.05 | 0.05 | 0.05 | 0.10 | 0.09 | 0.10 | 0.10 |

*Source:* Censo Escolar.

Table 7: **Municipality-level student flows by policy status**

| | Cycles | Grade 1 | Grade 2 | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|---|---|
| Repeat | | 0.16 | 0.14 | 0.10 | 0.09 | 0.13 | 0.11 | 0.08 | 0.07 |
| | No | 0.18 | 0.15 | 0.11 | 0.10 | 0.14 | 0.11 | 0.08 | 0.07 |
| | Yes | 0.04 | 0.08 | 0.04 | 0.09 | 0.07 | 0.07 | 0.05 | 0.08 |
| Pass | | 0.73 | 0.77 | 0.81 | 0.82 | 0.73 | 0.77 | 0.79 | 0.82 |
| | No | 0.70 | 0.75 | 0.79 | 0.81 | 0.71 | 0.76 | 0.78 | 0.82 |
| | Yes | 0.88 | 0.86 | 0.90 | 0.85 | 0.83 | 0.84 | 0.85 | 0.82 |
| Drop-out | | 0.08 | 0.06 | 0.06 | 0.06 | 0.12 | 0.11 | 0.11 | 0.10 |
| | No | 0.09 | 0.07 | 0.07 | 0.07 | 0.13 | 0.11 | 0.11 | 0.11 |
| | Yes | 0.04 | 0.03 | 0.03 | 0.03 | 0.08 | 0.08 | 0.08 | 0.09 |

*Source:* Censo Escolar.

Table 8: **Municipality-average grade attainment by age**

| Year | age 6 | age 7 | age 8 | age 9 | age 10 | age 11 | age 12 |
|---|---|---|---|---|---|---|---|
| 1999 | 1.02 | 1.10 | 1.60 | 2.24 | 2.95 | 3.61 | . |
| 2000 | 1.02 | 1.08 | 1.62 | 2.29 | 2.99 | 3.71 | 4.39 |
| 2001 | 1.03 | 1.09 | 1.66 | 2.36 | 3.08 | 3.79 | 4.54 |
| 2002 | 1.03 | 1.08 | 1.68 | 2.42 | 3.17 | 3.90 | 4.63 |
| 2003 | 1.02 | 1.09 | 1.70 | 2.46 | 3.24 | 4.00 | 4.75 |
| 2004 | 1.02 | 1.09 | 1.70 | 2.48 | 3.27 | 4.07 | 4.84 |
| 2005 | 1.02 | 1.09 | 1.68 | 2.47 | 3.28 | 4.10 | 4.90 |
| 2006 | . | 1.09 | 1.67 | 2.46 | 3.29 | 4.12 | 4.94 |

*Source: Censo Escolar*

Table 9: **Cycle prevalence: municipality means**

| **Year** | **Overall** | **Municipal** | **State** |
|---|---|---|---|
| 1999 | 0.192 | 0.209 | 0.421 |
| 2000 | 0.178 | 0.225 | 0.372 |
| 2001 | 0.174 | 0.223 | 0.376 |
| 2002 | 0.166 | 0.232 | 0.370 |
| 2003 | 0.160 | 0.229 | 0.356 |
| 2004 | 0.146 | 0.237 | 0.334 |
| 2005 | 0.189 | 0.238 | 0.406 |
| 2006 | 0.154 | 0.225 | 0.328 |

*Source: Censo Escolar.*

Table 10: **Movement in and out of cycle use**

| **Year** | **Change** | **Adopt** | **Unadopt** |
|---|---|---|---|
| 1999 | . | . | . |
| 2000 | 0.071 | 0.029 | 0.042 |
| 2001 | 0.048 | 0.022 | 0.026 |
| 2002 | 0.043 | 0.017 | 0.025 |
| 2003 | 0.035 | 0.015 | 0.021 |
| 2004 | 0.057 | 0.021 | 0.035 |
| 2005 | 0.075 | 0.059 | 0.016 |
| 2006 | 0.063 | 0.014 | 0.049 |

*Source: Censo Escolar.*

Table 11: **Municipality student flows (no lags): repeated**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | -0.0500*** | -0.0172*** | -0.0251*** | -0.00394*** | -0.0139*** | -0.00974*** | -0.0109*** | -0.00185 |
| | (0.000) | (0.000) | (0.000) | (0.004) | (0.000) | (0.000) | (0.000) | (0.185) |
| Observations | 38515 | 38512 | 38517 | 38518 | 38520 | 38513 | 38496 | 38460 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: **Municipality student flows: repeated**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | -0.0500*** | -0.0171*** | -0.0214*** | -0.00300 | -0.0215*** | 0.00430 | -0.00163 | 0.00834** |
| | (0.000) | (0.000) | (0.000) | (0.126) | (0.000) | (0.385) | (0.695) | (0.046) |
| L.Cycles | | 0.00497*** | -0.00528*** | -0.000607 | -0.00717** | 0.00751 | 0.000386 | 0.0215*** |
| | | (0.003) | (0.004) | (0.767) | (0.047) | (0.165) | (0.932) | (0.000) |
| L2.Cycles | | | 0.00654*** | 0.00663*** | 0.00365 | -0.000488 | 0.00293 | 0.00468 |
| | | | (0.000) | (0.002) | (0.379) | (0.943) | (0.607) | (0.414) |
| L3.Cycles | | | | 0.00320* | 0.00547 | 0.000686 | 0.00789 | 0.00722 |
| | | | | (0.093) | (0.139) | (0.912) | (0.132) | (0.170) |
| L4.Cycles | | | | | 0.0131*** | -0.000301 | 0.00161 | 0.000211 |
| | | | | | (0.000) | (0.958) | (0.739) | (0.965) |
| L5.Cycles | | | | | | -0.00416 | -0.00602 | -0.00885** |
| | | | | | | (0.359) | (0.115) | (0.021) |
| Observations | 38515 | 33016 | 27519 | 22017 | 16514 | 11011 | 11011 | 11011 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: **School-level student flows (no lags): repeated**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | -0.0838$^{***}$ | -0.0132$^{***}$ | -0.0298$^{***}$ | -0.00738$^{***}$ | -0.0158$^{***}$ | -0.0132$^{***}$ | -0.0136$^{***}$ | -0.00443$^{***}$ |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Observations | 1033310 | 1017858 | 990673 | 932982 | 353542 | 328867 | 310258 | 292118 |

*p*-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 14: **School-level student flows: repeated**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | -0.0838$^{***}$ | -0.0189$^{***}$ | -0.0335$^{***}$ | -0.0102$^{***}$ | -0.0147$^{***}$ | -0.00826$^{***}$ | -0.00620$^{***}$ | 0.00249 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.007) | (0.272) |
| L.Cycles | | 0.0177$^{***}$ | -0.000144 | 0.00339$^{**}$ | -0.00285$^{*}$ | 0.00205 | -0.00180 | 0.0107$^{***}$ |
| | | (0.000) | (0.912) | (0.027) | (0.077) | (0.412) | (0.428) | (0.000) |
| L2.Cycles | | | 0.00753$^{***}$ | 0.00614$^{***}$ | 0.00351$^{*}$ | 0.00492 | 0.00244 | 0.00186 |
| | | | (0.000) | (0.000) | (0.060) | (0.123) | (0.402) | (0.519) |
| L3.Cycles | | | | 0.00654$^{***}$ | 0.00137 | 0.00371 | -0.00258 | -0.000709 |
| | | | | (0.000) | (0.421) | (0.219) | (0.349) | (0.795) |
| L4.Cycles | | | | | 0.00963$^{***}$ | -0.000750 | -0.00224 | -0.00423$^{*}$ |
| | | | | | (0.000) | (0.788) | (0.379) | (0.094) |
| L5.Cycles | | | | | | -0.00716$^{***}$ | -0.00458$^{**}$ | -0.00616$^{***}$ |
| | | | | | | (0.003) | (0.034) | (0.004) |
| Observations | 1033310 | 839610 | 663388 | 495561 | 151321 | 95256 | 90923 | 85811 |

*p*-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 15: **Municipality student flows (no lags): promoted**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.0465$^{***}$ | 0.0104$^{**}$ | 0.0255$^{***}$ | 0.00276 | 0.0172$^{***}$ | 0.0149$^{***}$ | 0.0126$^{***}$ | 0.00755$^{*}$ |
| | (0.000) | (0.010) | (0.000) | (0.548) | (0.000) | (0.000) | (0.000) | (0.060) |
| Observations | 38515 | 38512 | 38517 | 38518 | 38520 | 38513 | 38496 | 38460 |

*p*-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 16: **Municipality student flows: promoted**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.0465*** | 0.00827* | 0.0196*** | 0.000496 | 0.0279*** | -0.00300 | -0.00276 | -0.00806 |
| | (0.000) | (0.094) | (0.001) | (0.955) | (0.001) | (0.841) | (0.819) | (0.535) |
| L.Cycles | | -0.00689 | 0.00462 | 0.00118 | 0.0105 | -0.00863 | -0.00243 | -0.0191 |
| | | (0.154) | (0.423) | (0.897) | (0.230) | (0.597) | (0.854) | (0.177) |
| L2.Cycles | | | -0.00825 | -0.00851 | 0.00578 | -0.00218 | -0.0108 | -0.0189 |
| | | | (0.141) | (0.376) | (0.566) | (0.915) | (0.511) | (0.288) |
| L3.Cycles | | | | -0.00674 | -0.00800 | -0.00475 | -0.0116 | -0.00186 |
| | | | | (0.428) | (0.374) | (0.801) | (0.444) | (0.909) |
| L4.Cycles | | | | | -0.00392 | -0.00110 | 0.000978 | -0.000562 |
| | | | | | (0.627) | (0.949) | (0.944) | (0.970) |
| L5.Cycles | | | | | | -0.00334 | -0.00510 | 0.00109 |
| | | | | | | (0.808) | (0.645) | (0.927) |
| Observations | 38515 | 33016 | 27519 | 22017 | 16514 | 11011 | 11011 | 11011 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 17: **School-level student flows (no lags): promoted**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.0713*** | 0.0220*** | 0.0211*** | 0.0191** | 0.00438*** | 0.0260*** | -0.00436*** | 0.0248*** |
| | (0.000) | (0.000) | (0.000) | (0.027) | (0.001) | (0.000) | (0.002) | (0.000) |
| Observations | 1033310 | 1017858 | 990673 | 932982 | 353542 | 328867 | 310258 | 292118 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 18: **School-level student flows: promoted**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.0713*** | 0.0282*** | 0.0212*** | 0.0386* | 0.00939*** | 0.00591 | -0.00276 | -0.00147 |
| | (0.000) | (0.000) | (0.000) | (0.050) | (0.001) | (0.231) | (0.526) | (0.778) |
| L.Cycles | | -0.0134*** | 0.00105 | 0.00311 | 0.0151*** | 0.00235 | 0.000263 | -0.0122** |
| | | (0.000) | (0.641) | (0.876) | (0.000) | (0.632) | (0.952) | (0.018) |
| L2.Cycles | | | -0.00325 | -0.0111 | 0.00722** | 0.00380 | -0.00606 | -0.0133** |
| | | | (0.135) | (0.578) | (0.032) | (0.545) | (0.273) | (0.045) |
| L3.Cycles | | | | -0.00888 | -0.0000692 | -0.0130** | 0.00732 | -0.00545 |
| | | | | (0.628) | (0.982) | (0.029) | (0.161) | (0.386) |
| L4.Cycles | | | | | -0.00980*** | -0.000738 | -0.00395 | 0.0111* |
| | | | | | (0.001) | (0.893) | (0.414) | (0.057) |
| L5.Cycles | | | | | | -0.00155 | -0.00271 | 0.00136 |
| | | | | | | (0.740) | (0.510) | (0.782) |
| Observations | 1033310 | 839610 | 663388 | 495561 | 151321 | 95256 | 90923 | 85811 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 19: **Municipality student flows (no lags): dropped**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | -0.00276** | -0.00100 | -0.00284*** | -0.00350*** | -0.00303* | -0.00313** | -0.00124 | -0.00132 |
| | (0.044) | (0.339) | (0.009) | (0.002) | (0.051) | (0.035) | (0.453) | (0.420) |
| Observations | 38515 | 38512 | 38517 | 38518 | 38520 | 38513 | 38496 | 38460 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 20: **Municipality student flows: dropped**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | -0.00276** | -0.000338 | 0.000581 | -0.000926 | -0.000961 | -0.00184 | -0.000912 | 0.000616 |
| | (0.044) | (0.775) | (0.634) | (0.420) | (0.675) | (0.586) | (0.790) | (0.880) |
| L.Cycles | | 0.000368 | -0.00165 | -0.000924 | -0.00295 | -0.00336 | -0.00153 | 0.000285 |
| | | (0.751) | (0.185) | (0.443) | (0.228) | (0.362) | (0.683) | (0.949) |
| L2.Cycles | | | -0.000112 | -0.0000782 | -0.00244 | -0.00487 | 0.00500 | 0.00663 |
| | | | (0.926) | (0.951) | (0.387) | (0.293) | (0.287) | (0.237) |
| L3.Cycles | | | | -0.00165 | 0.00161 | -0.000795 | 0.00133 | -0.00362 |
| | | | | (0.142) | (0.523) | (0.852) | (0.757) | (0.481) |
| L4.Cycles | | | | | 0.00137 | 0.000330 | 0.00123 | -0.00125 |
| | | | | | (0.543) | (0.933) | (0.757) | (0.793) |
| L5.Cycles | | | | | | 0.00637** | 0.00443 | -0.00173 |
| | | | | | | (0.040) | (0.160) | (0.645) |
| Observations | 38515 | 33016 | 27519 | 22017 | 16514 | 11011 | 11011 | 11011 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 21: **School-level student flows (no lags): dropped**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.00303*** | -0.00186** | 0.00541*** | -0.00238*** | 0.00847*** | -0.000110 | 0.0121*** | 0.00415*** |
| | (0.000) | (0.022) | (0.000) | (0.009) | (0.000) | (0.885) | (0.000) | (0.000) |
| Observations | 1033310 | 1017858 | 990673 | 932982 | 353542 | 328867 | 310258 | 292118 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 22: **School-level student flows: dropped**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.00303*** | -0.00177* | 0.00777*** | -0.000941 | 0.0151*** | 0.001000 | 0.000555 | 0.00377 |
| | (0.000) | (0.058) | (0.000) | (0.492) | (0.000) | (0.628) | (0.793) | (0.347) |
| L.Cycles | | -0.000145 | 0.000613 | -0.000546 | 0.00313** | -0.00215 | 0.00161 | 0.00534 |
| | | (0.874) | (0.571) | (0.694) | (0.030) | (0.296) | (0.445) | (0.178) |
| L2.Cycles | | | -0.00156 | 0.00153 | 0.00153 | -0.00613** | -0.00192 | -0.00434 |
| | | | (0.136) | (0.268) | (0.357) | (0.020) | (0.477) | (0.395) |
| L3.Cycles | | | | -0.00169 | -0.00142 | -0.00194 | -0.00711*** | 0.00152 |
| | | | | (0.185) | (0.349) | (0.435) | (0.005) | (0.752) |
| L4.Cycles | | | | | 0.000552 | 0.000718 | -0.00233 | -0.00472 |
| | | | | | (0.697) | (0.754) | (0.322) | (0.290) |
| L5.Cycles | | | | | | 0.00392** | 0.00458** | 0.000104 |
| | | | | | | (0.046) | (0.022) | (0.978) |
| Observations | 1033310 | 839610 | 663388 | 495561 | 151321 | 95256 | 90923 | 85811 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 23: **Municipalities: ln total enrollments (no lag)**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | -0.00729 | 0.00369 | -0.000989 | 0.00142 | -0.00474 | -0.00434 |
| | (0.208) | (0.277) | (0.769) | (0.652) | (0.123) | (0.181) |
| Observations | 44009 | 44027 | 44027 | 44027 | 44027 | 38531 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 24: **Municipalities: ln total enrollments**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | -0.00588 | 0.00442 | 0.00429 | 0.00187 | -0.0183*** | -0.00226 |
| | (0.373) | (0.288) | (0.369) | (0.716) | (0.004) | (0.706) |
| L.Cycles | -0.00824 | -0.00127 | -0.00742 | 0.0000907 | -0.00314 | -0.0119** |
| | (0.206) | (0.767) | (0.133) | (0.986) | (0.628) | (0.048) |
| L2.Cycles | | 0.00374 | -0.00589 | 0.00220 | 0.00603 | 0.00303 |
| | | (0.364) | (0.241) | (0.691) | (0.380) | (0.635) |
| L3.Cycles | | | 0.00440 | -0.00105 | 0.0101 | -0.00157 |
| | | | (0.373) | (0.857) | (0.199) | (0.829) |
| L4.Cycles | | | | -0.00356 | 0.0101 | -0.0000353 |
| | | | | (0.494) | (0.151) | (0.996) |
| L5.Cycles | | | | | -0.00351 | 0.00155 |
| | | | | | (0.576) | (0.790) |
| Observations | 38506 | 33023 | 27521 | 22017 | 16516 | 16516 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 25: **Schools: ln total enrollments (no lags)**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | 0.00546* | 0.00431 | 0.0132*** | 0.0466*** | 0.000598 | -0.0190*** |
| | (0.085) | (0.117) | (0.000) | (0.000) | (0.821) | (0.000) |
| Observations | 1147125 | 1208863 | 1221614 | 1253105 | 1251465 | 1017822 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 26: **Schools: ln total enrollments**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | 0.0000795 | 0.00102 | 0.00792** | 0.0235*** | 0.00633 | -0.0177*** |
| | (0.983) | (0.770) | (0.045) | (0.000) | (0.288) | (0.004) |
| L.Cycles | -0.00512 | 0.00876** | 0.00548 | 0.0147*** | -0.0138** | -0.0156*** |
| | (0.157) | (0.014) | (0.171) | (0.002) | (0.012) | (0.005) |
| L2.Cycles | | 0.00601* | 0.00857** | 0.0150*** | 0.00430 | -0.0133** |
| | | (0.077) | (0.029) | (0.001) | (0.457) | (0.024) |
| L3.Cycles | | | 0.0113*** | 0.0134*** | -0.00928 | -0.0217*** |
| | | | (0.003) | (0.005) | (0.140) | (0.001) |
| L4.Cycles | | | | 0.00758* | 0.000402 | -0.0138** |
| | | | | (0.084) | (0.944) | (0.017) |
| L5.Cycles | | | | | -0.00286 | 0.00103 |
| | | | | | (0.597) | (0.852) |
| Observations | 960116 | 831636 | 675264 | 539830 | 390857 | 366025 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 27: **Municipalities: average grade (no lag)**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | 0.00000424 | 0.0232*** | 0.0322*** | 0.0338*** | 0.0282*** | 0.0166*** |
| | (0.998) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| Observations | 44009 | 44027 | 44027 | 44027 | 44027 | 38531 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 28: **Municipalities: average grade**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | 0.00201 | 0.0233*** | 0.0294*** | 0.0300*** | 0.0306*** | 0.0154* |
| | (0.315) | (0.000) | (0.000) | (0.000) | (0.000) | (0.061) |
| L.Cycles | 0.000392 | 0.0240*** | 0.0279*** | 0.0295*** | 0.0327*** | 0.0350*** |
| | (0.843) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| L2.Cycles | | 0.00931*** | 0.0296*** | 0.0349*** | 0.0391*** | 0.0272*** |
| | | (0.007) | (0.000) | (0.000) | (0.000) | (0.002) |
| L3.Cycles | | | 0.00463 | 0.00964 | 0.0315*** | 0.0242** |
| | | | (0.317) | (0.118) | (0.001) | (0.016) |
| L4.Cycles | | | | 0.00606 | 0.0210** | 0.0154* |
| | | | | (0.269) | (0.013) | (0.086) |
| L5.Cycles | | | | | 0.00242 | 0.0106 |
| | | | | | (0.749) | (0.187) |
| Observations | 38513 | 33029 | 27526 | 22021 | 16519 | 16519 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 29: **Schools: average grade (no lags)**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | 0.000772 | 0.0343*** | 0.0283*** | 0.00535** | 0.00797*** | -0.0122*** |
| | (0.488) | (0.000) | (0.000) | (0.032) | (0.002) | (0.000) |
| Observations | 1147125 | 1208863 | 1221614 | 1253105 | 1251465 | 1017822 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 30: **Schools: average grade**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | 0.000638 | 0.0322*** | 0.0144*** | -0.00525 | -0.00323 | 0.00518 |
| | (0.626) | (0.000) | (0.000) | (0.226) | (0.586) | (0.434) |
| L.Cycles | 0.00513*** | 0.0254*** | 0.0414*** | 0.0244*** | 0.0330*** | 0.0199*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| L2.Cycles | | 0.00914*** | 0.0235*** | 0.0239*** | 0.0222*** | 0.0208*** |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) |
| L3.Cycles | | | 0.00216 | 0.0116*** | 0.0289*** | 0.0246*** |
| | | | (0.506) | (0.008) | (0.000) | (0.000) |
| L4.Cycles | | | | -0.00770* | 0.0130** | 0.0178*** |
| | | | | (0.058) | (0.022) | (0.005) |
| L5.Cycles | | | | | -0.000571 | -0.00433 |
| | | | | | (0.916) | (0.470) |
| Observations | 960116 | 831636 | 675264 | 539830 | 390857 | 366025 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

# A Data

## A.1 Enrollment

### A.1.1 Descriptive Statistics

Tables 31, 32 and 33 present municipality student enrollments in natural logs, levels and as a share of birth cohort.

### A.1.2 Enrolled Share

To calculate the share of a birth cohort that is enrolled at any point in time, we combine student data from the *Censo Escolar* with age-specific counts from the year 2000 Brazilian census. We do so by summing the counts of students of each age across grades, and dividing this total by the count of children of that same age from the census. While the census data allow us to have an external measure of total cohort size, several issues emerge.

First, the census reports ages of children in blocks, rather than by year. We have counts, specifically, of children aged five to six, seven to nine, and ten to fourteen. To deal with this, we use $(\frac{1}{m} * N)$ as the denominator for our enrolled share, where $m$ is the number of ages aggregated (for the count of five and six-year-olds, $m = 2$), and $N$ is the total count.[28]

The second issue is one of data availability. The youngest children for whom we have census data were five years old in 2000 (born in 1995), and therefore enter our analysis in 2001 when they turn six. For cohorts born after 1995, we cannot calculate the share of the cohort enrolled in school, because we do not have an estimate of cohort size which is exogenous to enrollment. Given that data on cycles begins in 1999, and that our primary estimation requires this data for years in which students are seven and eight years old, this constitutes a serious constraint.

Finally, although the approach outlined above should, in theory, provide an unbiased estimate of the share of a birth cohort enrolled at any given time, there are substantial disparities between the population counts reported in the *Censo Escolar* and in the census. As can be seen in Table 34,

[28] Note that this amounts to assuming that half of the sum of, e.g. five- and six-year-olds is an unbiased estimate of the number of five-year-olds or the number of six-year-olds.

enrollment figures are systematically larger than counts of children of the same age from the census, starting from age eight. The most likely explanation for this is over-reporting by schools of annual enrollment figures. Indeed, since the introduction of the *Fundo de Manutenção e Desenvolvimento do Ensino Fundamental* (FUNDEF) in 1998, transfers to municipalities for spending on primary education were tied to the number of enrolled students reported in the *Censo Escolar*. Further evidence for such an explanation comes from the fact that, when unique student identifiers were introduced the the *Censo Escolar* in 2007 - effectively making it more difficult to over-report - student numbers fell significantly.

We cannot test this theory directly, but we can compare municipalities which adopt cycles with those who do not. While we do not have an unbiased second estimate of cohort size, as a first check we can compare the maximal enrollment figure for a given cohort - that is, the largest number of students enrolled in any year - to our census cohort estimate. Table 35 compares the percent difference in enrollment of the 1994 cohort for municipalities with and without cycles. Note that this percent difference is constant across years (the census estimate and the maximal enrollment are time-invariant), however municipalities move between the two groups depending on their current cycles policy. While this test is in no way definitive, it gives us some confidence that the two groups of municipalities are not wildly different.

## A.2 Student Flows

Table 36 shows how passing and repetition rates have varied over time for grade 1 students.

## A.3 Control Variables

Summary statistics for the list of school covariates are given in Table 37, while municipal covariates are given in Table 38. All variables are taken directly from the *Censo Escolar* and IPEA, except for *Training of teachers* (at levels 1-4 and 5-8). This last variable is an index of mean education levels of teachers teaching at the specified grade levels, coded such that 0 represents less than primary education, 1 is completed primary education, 2 is completed secondary education, and 3 is any form of tertiary training. Summary statistics on the number and education levels of teachers in the school panel are calculated conditional on having at least one teacher teaching at that level.

# B The Cycles Policy

Figure 1 shows that the share of schools in a municipality which use cycles is highly bimodal. Note that the data are aggregated over all years in our sample. Most municipalities make no use of cycles; those that do, however, commonly adopt entirely. This bimodality is even more pronounced when looking at school jurisdictions individually, as can be seen in Figure 2.

## B.1 Adoption: Overall and by Dependency

Table 40 lists the share of municipalities coded as 'adopters' for varying thresholds. Tables 41 and 42 replicate this for municipal-run schools and state-run schools separately.

## B.2 Geographic Variation

The popularity of cycles policies varies considerably across regions. A brief description of the general patterns follows: these overviews are based on a visual inspection of the distribution of municipality-level adoption rates for the years 1999, 2001, 2003 and 2005, for state-run and municipality-run schools separately.

**Cycles in the North** Municipal schools in the North have low or zero cycle adoption rates over the period. Rates are similarly low in state schools, with a few exceptions: state schools in Roraima report some cycle use (with a few municipalities registering a 100% adoption), while state schools in Tocantins have a range of adoption rates in 1999, diminishing to zero by 2003.

**Cycles in the Northeast** Municipal schools in the Northeast have low or zero cycle adoption rates, with the exception of Rio Grande do Norte which displays a strong bimodal distribution of municipalities: some adopt at near-census rates, while others avoid the policy entirely (rates peak in 2001-2003). Cycles are more prevalent among state schools. While half of the states have low or zero adoption, Ceará and Bahia display a 'messy' bimodal distribution (with some interior mass) from 1999-2003, and Pernambuco has such an adoption pattern in 2005 only (with no cycles prior to this). State schools in Rio Grande do Norte have a messy bimodal adoption pattern in 1999, which strengthens to a strong level of adoption in 2005 (most municipalities at 100%, and no mass at zero).

**Cycles in the South** Both municipal schools and state schools in the south adopted cycles at trivial rates, with the stark exception of municipal schools in Paraná. Municipal schools in Paraná display a distinctly bimodal adoption rate: most municipalities either fully adopt, or do not adopt cycles at all.

**Cycles in the Southeast** Municipal schools in the Southeast display a strongly bimodal distribution of adoption rates (Espírito Santo deviates slightly from this trend in 2005, with more interior points). State schools in general all adopted cycles. Exceptions to this are Minas Gerais in 2001 and 2003, and Espírito Santo in 2005, which are bimodal.

**Cycles in the Centre-West** The Centre-West region does not seem to follow a common trend. In Goiás, no schools adopted cycles at any point. In Mato Grosso do Sul, from 1999-2003, state schools all had cycles, while municipal schools mostly didn't, with some exceptions (including several with 100% adoption). In 2005, these rates fall to zero in both dependencies. Both municipal and state schools in Mato Grosso display bimodal adoption rates throughout the time period, with non-trivial interior mass among state schools.

## B.3 Weighting

Our primary cycles variable is computed using the share of schools in a given municipality in a given year who report using cycles, weighted by the primary school enrollment of those schools in that year. Table 43 presents a summary statistic comparison between weighted and unweighted measures.

# C Robustness

## C.1 Total Enrollment in Levels

Tables 44 and 45 replicate the natural log enrollment regressions (see Tables 24 and 26 in Section 5.2), this time in levels rather than natural logs. A comparison of the means and standard deviations of these variables is given in Appendix A.1. The municipality regressions suggest that a history of cycles may have some positive impact on enrollment at ages eleven and twelve, while contemporaneous cycles have a negative effect at age eight. School-level regressions in Table 45 present much the same story as the equivalent table in natural logs: increases in enrollment at junior primary ages seven to ten, and decreases at age twelve. Note that the school-level regression are conditional on enrollments at that age being positive. This maintains the same sample as the natural log enrollment regression, and prevents schools which do not offer higher grades from displaying 'negative' effects on enrollment as they reduce age-for-grade mismatches.[29]

## C.2 Enrolled Share of Cohort

Increasing enrollment is in fact an indirect measure of a more fundamental goal: achieving universal enrollment of primary-school aged children. Using counts of children of different ages from the 2000 census, we replicate our municipality-level regressions using the enrolled share of a birth cohort. This process, and some of the issues which arise from the measure, are described in Appendix A.1.2.

Results from age-specific regressions with enrolled share of birth cohort as the dependant variable are presented in Table 46. There appears to be a permanent, negative effect of policy exposure at age seven, and a positive, similarly lasting, effect of cycles at age ten. In contrast to the coefficients found for enrollment in natural logs or in levels, the magnitude of these effects is significant: increases and decreases of 5-8 percent of the birth cohort.

While the direction of the effects is potentially plausible, we are sceptical of these results for several reasons. First, we fail to replicate these results in either natural logs or levels of actual

[29]Consider a primary school that offers only grades 1-4: unless children are very delayed in their schooling, there should be no twelve-year-olds at that school.

enrollment, even when including age-specific counts as control variables. Second, due to the requirement that children be five years old or older at the time of the census (in order to be counted), and yet have most of their schooling post-1999 (to include a history of the policy), very few birth cohorts are retained for analysis. Indeed, for early grades we have only the 1993-1995 cohorts (who were counted in the 2000 census, and are in the panel for both ages six and seven), while the estimates for twelve-year-olds come from the 1992-1994 cohorts. There is therefore a real concern that idiosyncratic effects of a single cohort could play a large role in these results.

## C.3 Alternate Cycles Measures

As described in Section 3.3, we aggregate individual schools' adoption of the cycles policy to create a municipal-level variable. In our primary specification, we code a municipality as having adopted the policy if at least 75% of all schools in the municipality (weighted by enrollment) have done so. While the bi-modality of policy adoption rates within municipalities suggests this is reasonable, we explore several other thresholds to be sure this is the case. Table 47 presents a series of regressions where the municipality repetition rate of grade 1 students is the dependant variable. In the regressions, the threshold for coding a municipality as having cycles varies from 10% of schools having adopted, to 90%. Table 40 in Appendix B gives the share of municipalities coded as using cycles for each of these thresholds.

Not surprisingly, the effect of cycle adoption on repetition rates falls as we weaken our definition of municipality-level adoption: fewer and fewer schools are implementing cycles. It is also slightly smaller when we raise the threshold to 90%. This is unexpected given that treatment is 'strengthened'; however, the number of municipalities using cycles fall from 15-19% (using a 75% threshold), to 10-15% (using a 90% threshold), therefore one third fewer municipalities are included in the treated group.

## C.4 Bolsa Escola

During our panel, the *Bolsa escola* (later *Bolsa familia*) conditional cash transfer program was rolled out across Brazil. Data on the presence of the program at a given school was first collected in 2001:

by 2004, nearly every school was responding positively (see Table 39 in Appendix A.3). Glewwe and Kassouf (2012) show that *Bolsa escola* increased enrollment and promotion rates and reduced dropout. If the adoption of the cycles policy is correlated with availability of the *Bolsa* program - for instance, if some municipalities are 'early adopters' - this could confound our estimates.

To explore this possibility, we re-estimate our enrollment, promotion and dropout equations, controlling for the presence of the program in that municipality.[30] As the survey simply asked wether *Bolsa escola* exists at the school, but not how many students were eligible or enrolled, we follow Glewwe and Kassouf (2012) by using presence of the program as a binary indicator. We aggregate this at the municipality level, with *Bolsa* equal to one if any schools report the program.

Because data on *Bolsa escola* were first collected in 2001, the sample on which we estimate these equations is smaller than our baseline sample. To compare estimates with and without the Bolsa control, we first re-estimate our primary specification using only the subsample of municipalities with a valid Bolsa observation. We then estimate the same equation, with the addition of the Bolsa dummy variable.

Tables 48 and 49 present the results on enrollment. Compared to our baseline specification (see Table 24), restricting the sample to the Bolsa years does change our estimates. While our baseline specification shows no effect of the cycles policy on the natural log of enrollment, Table 48 suggests that, from 2001, cycles decrease enrollment of seven-year-olds and eleven-year-olds, while increasing enrollment of ten-year-olds. When we compare these estimates to Table 49, however, we see that the addition of a dummy variable for *Bolsa escola* has only negligible effects on our estimated cycles coefficients.

Tables 50 and 51 repeat this approach for passing rates, while Table 52 and 53 do the same for dropouts. In both cases, we observe only the slightest changes in estimates when controlling for the *Bolsa* program.

[30]We do not replicate this procedure for schools. Given that students can sort themselves across schools in response to both the cycle policy and the Bolsa program[31], it would be difficult or impossible to interpret any differences that emerged.

## C.5 Minas Gerais

In order to make our findings comparable to those of Koppensteiner (2014), we replicate our repetition rate and dropout rate regressions, this time restricting our analysis to Minas Gerais. We do this both for all schools, and for state-run schools only, as this last is the sample Koppensteiner studies. In both cases we do not include lagged values of the policy variable.

Tables 54 and 55 display results from regressions with repetition rates and dropout rates as outcome variables, for the state school sample. For comparison with our main results, Tables 56 and 57 display results from similar regressions using all schools in Minas Gerais. For repetition rates, the coefficients on the policy variable are in both cases substantially larger than those in our country-wide regressions, although the sign and precision of the estimates is maintained. The effect of the policy on dropout rates is also somewhat larger than in our baseline specification, though the magnitude remains modest.

Table 31: **Municipality: ln total enrollments by age**

| Stats | Age 6 | Age 7 | Age 8 | Age 9 | Age 10 | Age 11 | Age 12 |
|---|---|---|---|---|---|---|---|
| Mean | 3.248 | 5.251 | 5.492 | 5.512 | 5.527 | 5.528 | 5.518 |
| SD | 1.697 | 1.169 | 1.146 | 1.143 | 1.138 | 1.136 | 1.135 |

*Source:* Censo Escolar

Table 32: **Municipality: total enrollments by age**

| Stats | Age 6 | Age 7 | Age 8 | Age 9 | Age 10 | Age 11 | Age 12 |
|---|---|---|---|---|---|---|---|
| Mean | 81.7 | 484.4 | 617.7 | 627.2 | 633.1 | 635.2 | 631.3 |
| SD | 398.0 | 2542.2 | 3267.7 | 3266.6 | 3257.9 | 3275.1 | 3252.1 |

*Source:* Censo Escolar

Table 33: **Municipality: share enrolled by age**

| Stats | Age 6 | Age 7 | Age 8 | Age 9 | Age 10 | Age 11 | Age 12 |
|---|---|---|---|---|---|---|---|
| mean | 0.2 | 0.9 | 1.0 | 1.1 | 1.1 | 1.0 | 1.0 |
| sd | 0.2 | 0.3 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |

*Source:* Censo Escolar

Table 34: **Enrolled share over time by birth cohort**

| Year | 1995 | 1994 | 1993 | 1992 | 1991 | 1990 | 1989 | 1988 |
|---|---|---|---|---|---|---|---|---|
| 1999 | . | . | 0.20 | 0.86 | 1.04 | 1.00 | 1.04 | 1.04 |
| 2000 | . | 0.16 | 0.87 | 1.02 | 1.04 | 1.00 | 1.02 | 1.03 |
| 2001 | 0.17 | 0.87 | 1.04 | 1.04 | 1.05 | 0.99 | 1.01 | . |
| 2002 | 0.91 | 1.06 | 1.06 | 1.05 | 1.05 | 0.99 | . | . |
| 2003 | 1.08 | 1.08 | 1.07 | 1.05 | 1.04 | . | . | . |
| 2004 | 1.09 | 1.08 | 1.06 | 1.04 | . | . | . | . |
| 2005 | 1.09 | 1.07 | 1.05 | . | . | . | . | . |
| 2006 | 1.08 | 1.05 | . | . | . | . | . | . |

*Source: Censo Escolar and 2000 census*

Table 35: **Percent difference between maximum enrollment and census for students born in 1994**

| | Cycles | | No cycles | |
|---|---|---|---|---|
| **Year** | **mean** | **sd** | **mean** | **sd** |
| 1999 | 0.167 | 0.643 | 0.166 | 0.285 |
| 2000 | 0.167 | 0.637 | 0.166 | 0.282 |
| 2001 | 0.146 | 0.234 | 0.172 | 0.433 |
| 2002 | 0.148 | 0.231 | 0.172 | 0.435 |
| 2003 | 0.147 | 0.228 | 0.172 | 0.435 |
| 2004 | 0.163 | 0.617 | 0.168 | 0.289 |
| 2005 | 0.165 | 0.618 | 0.167 | 0.289 |
| 2006 | 0.168 | 0.641 | 0.166 | 0.287 |

*Source: Censo Escolar*

Table 36: **Municipality-level average grade 1 student flows**

| Year | Pass | Drop | Repeat |
|---|---|---|---|
| 1999 | 0.714 | 0.109 | 0.154 |
| 2000 | 0.699 | 0.114 | 0.157 |
| 2001 | 0.729 | 0.086 | 0.157 |
| 2002 | 0.740 | 0.069 | 0.156 |
| 2003 | 0.731 | 0.067 | 0.158 |
| 2004 | 0.737 | 0.069 | 0.167 |
| 2005 | 0.752 | 0.058 | 0.164 |

*Source: Censo Escolar.*

Table 37: **Time-varying school controls**

| Variable | Mean | Std. Dev. | N |
|---|---|---|---|
| Urban | 0.415 | 0.493 | 1372731 |
| Municipal | 0.704 | 0.456 | 1372731 |
| State | 0.186 | 0.389 | 1372731 |
| Total teachers | 12.122 | 16.92 | 1372731 |
| Primary teachers | 9.242 | 12.143 | 1372731 |
| Teachers teaching 1-4 | 5.099 | 6.186 | 1281859 |
| Training of teachers 1-4 | 2.061 | 0.592 | 1281859 |
| Teachers teaching 5-8 | 14.941 | 11.453 | 428136 |
| Training of teachers 5-8 | 2.679 | 0.402 | 428136 |

Table 38: **Time-varying municipal controls**

| Variable | Mean | Std. Dev. | N |
|---|---|---|---|
| Schools: number | 31.092 | 66.018 | 44056 |
| Schools: urban | 12.907 | 59.471 | 44056 |
| Schools: municipal | 21.902 | 31.86 | 44056 |
| Schools: state | 5.784 | 20.134 | 44056 |
| Schools: private | 3.399 | 28.181 | 44056 |
| Population | 31874.608 | 190634.984 | 44056 |
| Ln population | 9.366 | 1.129 | 44056 |
| Ln municipal gdp | 10.566 | 1.392 | 44056 |
| Municipal gdp | 227680.017 | 2601262.8 | 44056 |
| Teachers teaching 1-4 | 148.103 | 631.03 | 44056 |
| Training of teachers 1-4 | 2.244 | 0.375 | 44055 |
| Teachers teaching 5-8 | 144.922 | 768.349 | 44056 |
| Training of teachers 5-8 | 2.654 | 0.334 | 44028 |

Table 39: **Share of municipalities with *Bolsa* students**

| year | mean |
|---|---|
| 1999 | . |
| 2000 | . |
| 2001 | 0.469 |
| 2002 | 0.961 |
| 2003 | 0.983 |
| 2004 | 0.995 |
| 2005 | 0.995 |
| 2006 | 0.999 |

*Source:* Censo Escolar

Table 40: **Share of municipalities adopting cycles, by threshold of use**

| | 10 % | 25% | 33% | 50% | 66% | 75% | 90% |
|---|---|---|---|---|---|---|---|
| 1999 | 0.502 | 0.441 | 0.408 | 0.313 | 0.226 | 0.192 | 0.151 |
| 2000 | 0.478 | 0.419 | 0.385 | 0.293 | 0.206 | 0.178 | 0.134 |
| 2001 | 0.472 | 0.416 | 0.381 | 0.287 | 0.201 | 0.174 | 0.131 |
| 2002 | 0.467 | 0.414 | 0.382 | 0.286 | 0.197 | 0.166 | 0.120 |
| 2003 | 0.445 | 0.391 | 0.364 | 0.265 | 0.186 | 0.160 | 0.115 |
| 2004 | 0.429 | 0.384 | 0.354 | 0.258 | 0.176 | 0.146 | 0.104 |
| 2005 | 0.424 | 0.387 | 0.368 | 0.284 | 0.216 | 0.189 | 0.149 |
| 2006 | 0.371 | 0.338 | 0.320 | 0.238 | 0.176 | 0.154 | 0.117 |

*Source:* Censo Escolar

Table 41: **Share of municipalities with more than X% of municipal schools using cycles**

| | 10 % | 25% | 33% | 50% | 66% | 75% | 90% |
|---|---|---|---|---|---|---|---|
| 1999 | 0.266 | 0.242 | 0.235 | 0.222 | 0.207 | 0.202 | 0.191 |
| 2000 | 0.277 | 0.256 | 0.248 | 0.237 | 0.228 | 0.222 | 0.210 |
| 2001 | 0.275 | 0.251 | 0.243 | 0.232 | 0.220 | 0.214 | 0.203 |
| 2002 | 0.285 | 0.260 | 0.250 | 0.237 | 0.225 | 0.220 | 0.208 |
| 2003 | 0.276 | 0.250 | 0.244 | 0.234 | 0.222 | 0.218 | 0.207 |
| 2004 | 0.272 | 0.257 | 0.253 | 0.246 | 0.235 | 0.232 | 0.219 |
| 2005 | 0.278 | 0.260 | 0.256 | 0.246 | 0.235 | 0.231 | 0.219 |
| 2006 | 0.258 | 0.243 | 0.240 | 0.233 | 0.223 | 0.218 | 0.207 |

*Source: Censo Escolar*

Table 42: **Share of municipalities with more than X% of state schools using cycles**

| | 10 % | 25% | 33% | 50% | 66% | 75% | 90% |
|---|---|---|---|---|---|---|---|
| 1999 | 0.515 | 0.505 | 0.495 | 0.472 | 0.436 | 0.421 | 0.397 |
| 2000 | 0.484 | 0.475 | 0.467 | 0.446 | 0.399 | 0.380 | 0.343 |
| 2001 | 0.471 | 0.462 | 0.455 | 0.438 | 0.396 | 0.375 | 0.341 |
| 2002 | 0.462 | 0.450 | 0.444 | 0.428 | 0.390 | 0.366 | 0.328 |
| 2003 | 0.447 | 0.439 | 0.432 | 0.414 | 0.375 | 0.358 | 0.319 |
| 2004 | 0.430 | 0.422 | 0.417 | 0.402 | 0.366 | 0.345 | 0.303 |
| 2005 | 0.447 | 0.442 | 0.438 | 0.429 | 0.413 | 0.404 | 0.388 |
| 2006 | 0.392 | 0.386 | 0.381 | 0.371 | 0.343 | 0.329 | 0.303 |

*Source: Censo Escolar*

Table 43: **Cycle prevalence: comparison of weighted and unweighted measures**

| | Weighted | | | Unweighted | | |
|---|---|---|---|---|---|---|
| **Year** | **Overall** | **Municipal** | **State** | **Overall** | **Municipal** | **State** |
| 1999 | 0.317 | 0.212 | 0.423 | 0.242 | 0.190 | 0.416 |
| 2000 | 0.298 | 0.229 | 0.376 | 0.243 | 0.208 | 0.375 |
| 2001 | 0.293 | 0.233 | 0.372 | 0.240 | 0.212 | 0.372 |
| 2002 | 0.289 | 0.246 | 0.360 | 0.242 | 0.224 | 0.360 |
| 2003 | 0.274 | 0.239 | 0.338 | 0.234 | 0.220 | 0.340 |
| 2004 | 0.264 | 0.238 | 0.313 | 0.239 | 0.228 | 0.321 |
| 2005 | 0.286 | 0.241 | 0.364 | 0.248 | 0.232 | 0.364 |
| 2006 | 0.243 | 0.225 | 0.283 | 0.219 | 0.218 | 0.286 |

*Source: Censo Escolar.*

Table 44: **Municipalities: total enrollments**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Age 7 | Age 8 | Age 9 | Age 10 | Age 11 | Age 12 |
| Cycles | -5.675 | -4.672$^{*}$ | -2.061 | -2.596 | -4.258 | 0.885 |
| | (0.515) | (0.089) | (0.410) | (0.317) | (0.146) | (0.797) |
| L.Cycles | -2.942 | 0.131 | -0.786 | 0.575 | -1.739 | -0.910 |
| | (0.733) | (0.963) | (0.761) | (0.829) | (0.555) | (0.792) |
| L2.Cycles | | 2.724 | -1.417 | 4.011 | 6.424$^{**}$ | -1.951 |
| | | (0.317) | (0.589) | (0.151) | (0.040) | (0.594) |
| L3.Cycles | | | 1.433 | 1.981 | 10.38$^{***}$ | 4.301 |
| | | | (0.579) | (0.501) | (0.004) | (0.304) |
| L4.Cycles | | | | 1.343 | 7.086$^{**}$ | 6.744$^{*}$ |
| | | | | (0.609) | (0.026) | (0.072) |
| L5.Cycles | | | | | -2.230 | 3.235 |
| | | | | | (0.435) | (0.335) |
| Observations | 38524 | 33023 | 27521 | 22017 | 16516 | 16516 |

$p$-values in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 45: **Schools: total enrollments (conditional on positive enrollment at that age)**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | 0.357*** | 0.00423 | 0.186** | 0.446*** | 0.113 | -0.207 |
| | (0.000) | (0.953) | (0.019) | (0.000) | (0.376) | (0.112) |
| L.Cycles | -0.109 | 0.197*** | 0.0859 | 0.256*** | -0.318*** | -0.251** |
| | (0.128) | (0.007) | (0.285) | (0.004) | (0.007) | (0.035) |
| L2.Cycles | | 0.0110 | 0.0439 | 0.141 | 0.245** | 0.0298 |
| | | (0.875) | (0.578) | (0.115) | (0.048) | (0.811) |
| L3.Cycles | | | 0.248*** | 0.122 | -0.152 | -0.172 |
| | | | (0.001) | (0.175) | (0.259) | (0.209) |
| L4.Cycles | | | | 0.169** | -0.0121 | -0.321*** |
| | | | | (0.042) | (0.921) | (0.009) |
| L5.Cycles | | | | | 0.0600 | 0.242** |
| | | | | | (0.604) | (0.040) |
| Observations | 960116 | 831636 | 675264 | 539830 | 390857 | 366025 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 46: **Municipalities: share of cohort enrolled**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | -0.0195 | 0.0132 | -0.00290 | 0.0363*** | -0.0263** | 0.00461 |
| | (0.123) | (0.340) | (0.813) | (0.003) | (0.025) | (0.716) |
| L.Cycles | 0.00167 | -0.0237* | 0.00734 | 0.00280 | 0.0260** | -0.00731 |
| | (0.881) | (0.054) | (0.600) | (0.829) | (0.027) | (0.567) |
| L2.Cycles | | 0.00184 | -0.0299** | 0.0124 | 0.00423 | 0.0503*** |
| | | (0.867) | (0.018) | (0.407) | (0.735) | (0.000) |
| L3.Cycles | | | 0.00143 | -0.0315** | 0.0177 | -0.00452 |
| | | | (0.898) | (0.018) | (0.217) | (0.770) |
| L4.Cycles | | | | 0.00512 | -0.0129 | -0.0108 |
| | | | | (0.669) | (0.311) | (0.434) |
| L5.Cycles | | | | | 0.00419 | -0.0241* |
| | | | | | (0.714) | (0.051) |
| Observations | 16507 | 16507 | 16510 | 16511 | 16516 | 16516 |

$p$-values in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 47: **Municipality grade 1 repetition rates: alternate cycles measures**

| | (1) 10% | (2) 25% | (3) 33% | (4) 50% | (5) 66% | (6) 75% | (7) 90% |
|---|---|---|---|---|---|---|---|
| Cycles | -0.0315*** | -0.0346*** | -0.0383*** | -0.0412*** | -0.0480*** | -0.0500*** | -0.0451*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Observations | 38515 | 38515 | 38515 | 38515 | 38515 | 38515 | 38515 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 48: **Municipalities: ln total enrollments (no lag) - Bolsa sample**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | -0.0125* | 0.00412 | 0.00437 | 0.00654* | -0.00866** | -0.00580 |
| | (0.085) | (0.313) | (0.289) | (0.087) | (0.021) | (0.105) |
| Observations | 33012 | 33029 | 33029 | 33029 | 33029 | 33029 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 49: **Municipalities: ln total enrollments (no lag)**

| | (1) Age 7 | (2) Age 8 | (3) Age 9 | (4) Age 10 | (5) Age 11 | (6) Age 12 |
|---|---|---|---|---|---|---|
| Cycles | -0.0126* | 0.00410 | 0.00435 | 0.00653* | -0.00867** | -0.00581 |
| | (0.084) | (0.314) | (0.292) | (0.087) | (0.021) | (0.104) |
| Bolsa | 0.00960 | 0.00753** | 0.00995*** | 0.00645* | 0.00351 | 0.00661** |
| | (0.127) | (0.033) | (0.005) | (0.051) | (0.280) | (0.032) |
| Observations | 33012 | 33029 | 33029 | 33029 | 33029 | 33029 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 50: **Municipality student flows (no lags) - Bolsa sample: promoted**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.0412*** | 0.00315 | 0.0211*** | -0.0000710 | 0.0263*** | 0.0221*** | 0.0205*** | 0.0157*** |
| | (0.000) | (0.593) | (0.000) | (0.992) | (0.000) | (0.000) | (0.000) | (0.005) |
| Observations | 27517 | 27514 | 27519 | 27520 | 27522 | 27520 | 27515 | 27509 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 51: **Municipality student flows (no lags): promoted**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.0412*** | 0.00313 | 0.0211*** | -0.0000894 | 0.0263*** | 0.0221*** | 0.0205*** | 0.0157*** |
| | (0.000) | (0.595) | (0.000) | (0.990) | (0.000) | (0.000) | (0.000) | (0.005) |
| bolsa | 0.00266 | 0.00424 | 0.00465 | 0.00408 | 0.00329 | 0.00311 | 0.00452 | 0.00567 |
| | (0.566) | (0.344) | (0.271) | (0.431) | (0.383) | (0.440) | (0.202) | (0.187) |
| Observations | 27517 | 27514 | 27519 | 27520 | 27522 | 27520 | 27515 | 27509 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 52: **Municipality student flows (no lags) - Bolsa sample: dropped**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.000604 | 0.00190* | 0.000400 | -0.00112 | -0.00215 | -0.00383** | -0.000998 | -0.00691*** |
| | (0.708) | (0.098) | (0.740) | (0.355) | (0.221) | (0.022) | (0.578) | (0.000) |
| Observations | 27517 | 27514 | 27519 | 27520 | 27522 | 27520 | 27515 | 27509 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 53: **Municipality student flows (no lags): dropped**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 | (5) Grade 5 | (6) Grade 6 | (7) Grade 7 | (8) Grade 8 |
|---|---|---|---|---|---|---|---|---|
| Cycles | 0.000606 | 0.00190* | 0.000409 | -0.00112 | -0.00215 | -0.00383** | -0.000996 | -0.00689*** |
| | (0.707) | (0.097) | (0.734) | (0.355) | (0.221) | (0.022) | (0.579) | (0.000) |
| bolsa | -0.000576 | -0.00104 | -0.00196** | -0.0000574 | -0.0000331 | -0.000287 | -0.000542 | -0.00462*** |
| | (0.639) | (0.232) | (0.032) | (0.950) | (0.980) | (0.821) | (0.692) | (0.001) |
| Observations | 27517 | 27514 | 27519 | 27520 | 27522 | 27520 | 27515 | 27509 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 54: **School-level student flows in Minas Gerais state schools (no lags): repeated**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 |
|---|---|---|---|---|
| Cycles | -0.172*** | -0.115*** | -0.0968*** | -0.0970*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Observations | 16884 | 17012 | 17206 | 17388 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 55: **School-level student flows in Minas Gerais state schools (no lags): dropped**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 |
|---|---|---|---|---|
| Cycles | -0.00421* | -0.00493** | -0.00249 | -0.00364 |
| | (0.099) | (0.022) | (0.212) | (0.109) |
| Observations | 16884 | 17012 | 17206 | 17388 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 56: **School-level student flows in Minas Gerais (no lags): repeated**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 |
|---|---|---|---|---|
| Cycles | -0.163*** | -0.0729*** | -0.0807*** | -0.0315*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Observations | 81411 | 80995 | 80426 | 78806 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 57: **School-level student flows in Minas Gerais (no lags): dropped**

| | (1) Grade 1 | (2) Grade 2 | (3) Grade 3 | (4) Grade 4 |
|---|---|---|---|---|
| Cycles | -0.00546*** | -0.00560*** | -0.00839*** | -0.00503*** |
| | (0.001) | (0.000) | (0.000) | (0.002) |
| Observations | 81411 | 80995 | 80426 | 78806 |

$p$-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

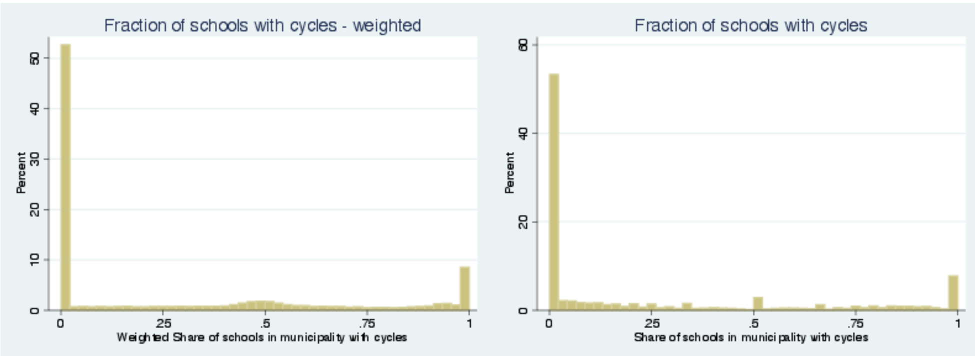



Figure 1: **Distribution of cycle frequency: municipal means**

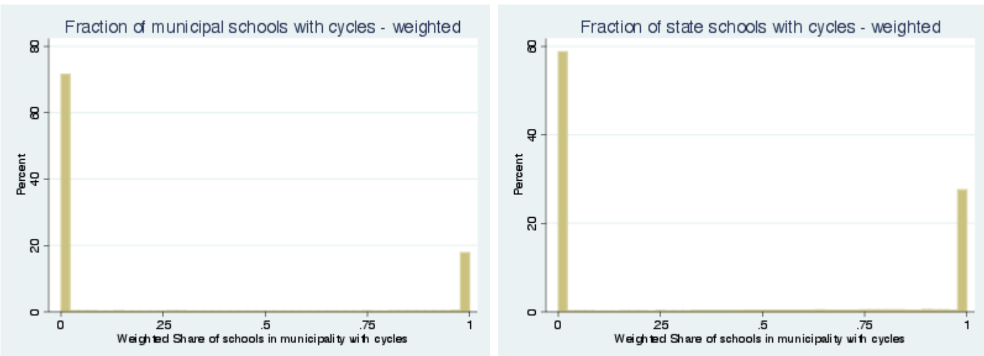



Figure 2: **Distribution of cycles: municipal vs. state schools (weighted)**

# Municipal Finance and Educational Investments: Evidence from Unexpected Budget Shocks in Brazil*

Julia Bird† and Margaret Leighton‡

May 25, 2015

**Abstract**

This paper exploits exogenous variation in the federal funds received by municipal governments in Brazil to examine the impact of transfers on local government revenues and expenditures, and in turn, on education provision. Consistent with a strong Flypaper Effect, we find that increased transfers lead to an immediate rise in current and capital spending. These increases are focused on education and welfare expenditure in poorer municipalities, while richer municipalities expand capital spending in the transport and housing sectors. Furthermore, particularly in wealthier municipalities, increases in transfers cause a short-term increase in local tax revenues. Evidence from municipal education resources suggests that these increases in spending are not being wasted. Positive transfer shocks are associated with increases in the number of teachers and, to a lesser extent, the number of classrooms. Transfers are also associated with substantial re-allocation of resources across schools offering classes at different levels, with secondary schools and schools teaching senior primary grades expanding at the expense of junior primary schools.

JEL classification H21; H71; H77; I2; R51

Keywords Fiscal federalism; Grants; Local taxation; Local government expenditure; Education finance

*We thank Stéphane Straub, Marie-Françoise Calmette, David Childers, Marie Lalanne, Mattia Girotti and Michel Azulai for insightful discussions and comments.

†Toulouse School of Economics, juliahbird@gmail.com.

‡Toulouse School of Economics, mleighton@gmail.com.

# 1 Introduction

Governments around the world frequently use inter-regional transfers to target disadvantaged populations and to finance improvements in their access to services. The use of international transfers to help those living in poor countries has sparked vigorous debates, with many arguing that transfers do little to help, and that they may in fact hinder economic development.[1] Many of these same arguments can apply to transfers between regions within the same country.

First, transfers that are not established from local tax revenues may be more exposed to elite capture: if the local population is poorly informed about the existence and size of these funds, the local government is less accountable for how they are used. Second, some portion of these transfers may be wasted, as looser budget constraints reduce the incentives for local governments to spend wisely. Finally, transfers may crowd out local revenue-raising, allowing the local government to slacken tax collection efforts, and thereby lower local revenues in the future.

This paper evaluates the impact of federal transfers to municipal governments on the provision of education in Brazil. We exploit the fact that federal transfers are distributed within states using a discontinuous population-based sharing formula in order to identify exogenous changes in municipal transfers. Specifically, we instrument the size of transfers by using the population of other municipalities within the same state. While population estimates of a given municipality are potentially open to manipulation, or may even induce municipalities to subdivide, the population of other municipalities within the same state is exogenous. We then use these exogenous changes to study the impact of transfers on spending and revenue, and on the number of classrooms and teachers in the municipality.

A substantial literature has examined the effect of transfers on local government expenditures and revenue generation. Traditionally, theory had predicted that transfers from national governments to local subsidiaries would crowd out funds from local revenue sources. Local governments already optimise the share of public and private spending within their municipality. An increase in revenues from external sources will lead to a less-than-proportionate rise in local

[1]See, for example, the recurring debates between William Easterly, Dambisa Moyo and Jeffrey Sachs (among others), and the books which have become synonymous with this issue: *The White Man's Burden* (Easterly [14]), *The End of Poverty* (Sachs [32]) and *Dead Aid* (Moyo [28]).

government spending, as the government with 'return' some of the transfer to local taxpayers through a reduction in their tax rates (Bradford and Oates, 1971 [9]).

Empirical evidence, however, suggests that this happens to a far lesser degree than such a simple model would predict. Hines and Thaler (1995 [20]) provide an early overview of the subject. They document cases in which, when the national government throws money at local governments through a transfer, it sticks where it lands: the so-called Flypaper Effect. There are various plausible reasons for this absence of crowd-out. First, local taxation results in a deadweight loss. External funds, which do not impose deadweight loss at the local level, may have a higher propensity to be spent than local revenues. Increases in external funds therefore may not result in lower local taxes. Secondly, voters may view an inflow of outside funds differently than they view local resources. While voters may not support a tax increase to expand public expenditure, when the funds arrive from elsewhere they are more than willing to see them spent on public projects.[2]

It is only recently that this literature has adopted a focus on exogenous variation in transfers. Dahlberg et al. (2008 [12]) use a migration-based discontinuity in the level of transfers received by municipalities in Sweden between 1996 and 2004 to study the impact of transfers on local government finance. They find evidence supportive of a Flypaper Effect: local taxes are unaffected by the increase in transfers, and local spending rises. Gordon (2004 [18]) looks at educational transfers in the US. She exploits discontinuous changes in the funding received by school districts which result from block grant allocations being updated after each decennial census. She finds that grant dollars translate directly into increased spending initially, but that within three years the new funding has crowded out other resources.

This paper is not the first to study variations in municipal transfers in Brazil. Gardner (2013 [17], chapter 2) explicitly addresses the extent to which federal transfers crowd out local revenue generation. Using variation in the laws which govern the repartition of Brazil's value-added tax in the Northeast states of Rio Grande do Norte and Paraíba, she finds no evidence for crowd-out. Several papers exploit the population-based discontinuity in the *Fundo de Participacao*

[2]Hines and Thaler note that, if such a behavioural effect were important, it should not be limited to the public sector. Indeed, a similar effect has been observed in the private sector. In that case, a surprise increase in cash flow to firms has been found to flow more readily towards investment than towards dividends to shareholders, even when the profitability of such investments is no better than it was before (see Blanchard et al., 1994[7]).

*dos Municipios* to study the effect of transfers on education. Using large discontinuous funding changes in the 1980s, Litschig & Morrison (2013 [25]) find that increases in transfers have long-lasting effects on children's school achievement. Bastos & Straume (2013 [4]) find a positive effect of municipal transfers on the public provision of preschool; furthermore, this increase does not crowd out private provision. In contrast, Gadenne (2014 [16]) finds that, while exogenous increases in tax revenues do increase municipal school infrastructure, increases in transfers have no effect.

Our paper makes three contributions to this literature. First, using the universe of Brazilian municipalities, we show that transfers do not crowd out local tax-raising efforts. In wealthier municipalities, an increase federal transfers may in fact *raise* tax revenues, resulting in higher and enduring capital spending. Second, increases in transfers lead to temporary increases in current spending. Poorer municipalities increase their spending particularly in welfare and education, whereas richer municipalities have higher capital spending increases in transport and in housing and urbanism. These findings together give strong evidence for the presence of a Flypaper Effect.

Finally, we are able to link these spending changes to genuine changes in the resources available for education. By doing so, we show that these increases in spending are not ineffective: measurable inputs to the education system increase immediately with transfers. Positive transfer shocks are associated with increases in the number of teachers and, to a lesser extent, the number of classrooms. For example, a 10% increase in the principal federal transfer leads to a 0.6% increase in the number of teachers, with larger increases in poorer municipalities. Furthermore, increases in transfers are associated with substantial re-allocation of resources across schools teaching at different levels, with schools offering more advanced – and possibly more expensive – grade levels expanding at the expense of lower primary schools.

The remainder of paper is structured as follows. Section 2 provides context on government spending, revenues and powers within Brazil, as well as background information on the education system. Section 3 outlines the data, while Section 4 details our estimation strategy. Sections 5 and 5.4 present our main results, Section 6 discusses them in relation to previous papers, and Section 7 concludes.

## 2 Municipal Public Finance in Brazil

### 2.1 Overview of Municipal Transfers

The three-tiered government of Brazil is one of the most decentralised in the world. Following the 1988 constitution, which entered into force after years of military rule, each of the three layers of government were accorded clearly-defined funding sources and obligations. Over 45% of all government expenditures within Brazil are spent at the municipal level, and yet only a third of their revenues are raised locally through taxes. The remaining 67% come from state and federal transfers.[3] Figure 1 shows the average share of resources from different sources in municipal government budgets in 2005.

While in 1970 there were 3952 municipalities in Brazil, that number has nearly doubled to 5564 in 2010. The average municipal population is around 34 000 people; however, municipalities range in size between the smallest, counting just 239 inhabitants, to the largest, São Paolo, with a population of 3.5 million. Municipalities are based around a single urban area: the administrative centre from which they also take their name. Over time, as municipalities have grown, the emergence of a large secondary town has often lead to municipalities splitting. According to the constitution, municipalities have the right to divide or merge if such a measure is supported by popular vote. This has meant that many municipalities with two major towns have preferred to divide in order to give both towns more autonomy. Figure 2 shows the municipalities of Brazil in 2000, according to their GDP per capita. As can been seen, there is considerable variation in GDP per capita across Brazil, both across and within states. In particular, the southern half of the country is far richer than the north.

Each of the three levels of government is responsible for applying different taxes, documented in Table 1. For the majority of these taxes, obligations are in place which require the government to transfer a share of the revenues to other levels of government. Due to the highly decentralised structure of Brazil, the richer states in the South and Southeast of the country have higher potential revenues and expenditure capacity than poorer states in other regions. Revenue-sharing in the form of federal and state transfers has been introduced to try and reduce such disparities.

[3]These revenue-raising shares are from 2005. See: World Bank Fiscal Decentralization Indicators, http://www.worldbank.org/publicfinance/decentralization, and Tresauro Nacional http://www.tesouro.fazenda.gov.br/ .

The allocation of funds across states and municipalities is defined in the constitution and hence fairly rigid over time. As such, the federal funds available to states and municipalities are largely without political discretion. The *Fundo de Participacao dos Municipios* (FPM), which will be the focus of this study, is one such formula-based transfer. One of the most important sources of municipal revenue, FPM transfers represented 42% of local government budgets in 2000, although the exact share varies considerably from year to year.[4] The federal government is required to transfer 22.5% of total revenues from income tax and industrial VAT (IPI and IR, respectively; two of the highest yielding federal taxes) through this fund, to be distributed to municipalities according to a predetermined formula. We will return to this formula, which is based around a municipality's population and that of other municipalities within the same state, in Section 4.1. Besides a percentage of these transfers which is earmarked for health and education, the municipalities are free to spend the money as they wish.

Municipalities hold independent executive and legislative power, with a directly-elected mayor (*Prefeito*) and an elected legislative body (*Camera dos Vereadores*). Elections are held every four years, with the last elections occurring in 2012. Each municipality is free to decide their own laws, within state and federal limits, and impose their own taxes, as well as receiving transfers from the state and federal governments. They are also responsible for particular services. State governments are also elected four-yearly, with residents directly electing a governor and a legislative council. State-level elections occur at the same time as the presidential election, offset by two years from the municipal election cycle.

Municipal governments have four principal taxes at their disposal: the property tax on urban buildings and land[5] (IPTU), a tax on local services (ISS), and taxes on imported services and transfers of real estate. The former two taxes are the principal, non-transfer sources of municipal income. While the size of the taxes may vary across municipalities, some restrictions apply: the ISS, for example, must fall between 2% and 5%.

While the constitution clearly specifies the revenue sources for each layer of government, it is less explicit when it comes to expenditures. For housing and sanitation, public transport, and

[4]This average excludes municipalities larger than 156,216 people, which are typically richer municipalities with larger tax revenues.

[5]Paid yearly by the owner of a plot of urban land or a building; rural land taxes are under the jurisdiction of the federal government.

the environment and natural resources, for example, it is not clear which level of government is responsible for provision. In addition, due to disparities in the institutional capacity at different levels of government – particularly immediately following the adoption of the constitution – states and the federal government were reluctant to pass down the responsibility for important services to municipal governments, despite the fact that these local authorities were receiving an increased share of revenues (Afonso & de Mello (2000)[1]).

## 2.2 Education Finance: the FUNDEF

Primary education in Brazil spans the ages of 7 to 14. Those in junior primary grades[6] 1-4 are taught by a single teacher, while senior primary grades 5-8 are taught by subject-specific teachers. This is followed by secondary education which lasts for three years. Enrollment rates at both primary and secondary levels have increased remarkably over the past two decades. Indeed, it has been noted that in Brazil, the "rise in the average educational attainment of the labour force since 1995 has been one of the fastest on record and faster than China's, which had been a global leader in schooling expansion in the prior decades. (Bruns et al. (2010)[11], p.3)" While fewer than half of all children completed primary school in 1995, in the nine years between PISA 2003 and PISA 2012, the number of 15-year-olds enrolled in school increased from 65% to 78% (OECD (2013)[30]).

All teachers in Brazil are required to have at minimum a secondary-level education, with a higher level required for grades 5-8 and up. In reality this is not always the case: in 1999, 9.6% of junior primary school teachers had primary education or less, and only 23.3% had completed higher education. To remedy this, the government set a goal to ensure all primary school teachers have tertiary training by 2007 (de Castro (2001)[13] pp.74-76). By 2003, the proportion of under-qualified teachers in lower primary school had fallen to 2%, and 61% of all teachers had tertiary education (Souza (2004)[22], p. 18).

Education finance has also been affected by the decentralisation movement of the past two decades. Passed in 1996, the *Lei de Diretrizes e Bases da Educação* (National Education Guidelines and Framework Law), in addition to decentralising additional authority to schools themselves,

[6]We use the word 'grade' as an analogy to the Brazilian term *serie*, with corresponding primary school levels 1-8.

clarified the duties of different levels of government with respect to education provision. Specifically, the law formally attributed responsibility for early childhood and primary education to the municipalities, while responsibility for secondary education was passed to state governments. The federal government, in turn, was given responsability for higher education. Despite this formal devolution of responsabilities, in reality the provision of most levels of education is still shared between municipalities and states.[7]

The 1996 law also introduced additional rules on education funding, and created an inter-regional funding-equalisation vehicle, the *Fundo para Manutenção e Desenvolvimento do Ensino Fundamental e Valorização do Magistério* (Fund for Maintenance and Development of Primary Education and Teacher Enhancement, FUNDEF). These finance reforms came into effect in 1998, and the FUNDEF was created.[8] The reforms included significant new rules determining minimal expenditures on education, and imposed constraints on some of those expenses. The law stipulates, for example, that a minimum of 25% of state and municipal revenues, and 18% of federal revenues, must be spent on education. Furthermore, 60% of the state and municipal minimum, and 30% of the federal minimum, must be spent on primary education (Levačić & Downes (2004) [23]).

These minimum expenditures are collected in a state-level fund, and redistributed to municipalities based on the number of students enrolled in primary school the previous year (Gardner (2013) [17], chapter 2). The law also stipulates minimum per-student amounts (updated annually): in the event that state-level amounts fall below this threshold, the federal government makes up the difference. Note that contributions to the FUNDEF are fixed share of revenues at each level of administration: the FUNDEF therefore acted as a further mechanism to redistribute resources towards poorer states. The introduction of the fund also had significant impacts on resource allocation within states.

A number of studies have been done to evaluate the impact of the FUNDEF reform. Gordon & Vegas (2004) [19] find increases in enrollment in states which benefited most from the reform,

[7]State-run primary schools were gradually, but not entirely, transferred to municipalities. Madeira (2012)[26] describes this decentralisation and devolution as it was carried out in Sao Paolo. The 1998 education finance reforms also formalized the the funding implication of such transfers: while arrangements had previously been on a case-by-case basis, from this point on the transfer of schools from state to municipality was associated with a strictly defined transfer of resources.

[8]See Gardner (2013 [17], chapter 1) for a detailed description of municipal revenue flows and expenditures, including the FUNDEF (and successor from 2007, FUNDEB).

and that funding increases were often used to decrease class sizes. Menezes-Filho & Pazello (2007) [27] find that salary increases due to the reform are associated with an increase in proficiency of grade 8 students. Estevan (2014) [15] uses the FUNDEF to study to what extent improvements in public schools draw students out of private institutions. She finds that quality improvements resulting from the reform increased the share of grade 1 students enrolled in public schools significantly, and – to a lesser extent – of students in grades 2 and 4 as well.

# 3 Data

We merge data from several publicly-available administrative databases to create an 11-year panel, with municipality-years as the unit of observation. The panel is unbalanced, with 4840 municipalities in 1996 and 5400 in 2006. While we maintain municipalities which divide during our panel, we drop those that do not receive funding in accordance with the FPM formula described in Section 4. These excluded municipalities are those with a population larger than 156,216 and state capitals. We rely on *Instituto de Pesquisa Economica Applicada* (IPEA) for a the majority of our data, while data on education infrastructure and teachers comes from the *Censo Escolar,* an annual school census.

## 3.1 IPEA

The three censuses of 1991, 2000 and 2010, provide municipal-level population. In the intervening years, municipal population comes from estimates calculated by the *Instituto Brasileiro de Geografia e Estatística* (IBGE), also reported by IPEA. The finance variables, including FPM transfers, tax revenues, current and capital expenditure and other expenditures by function, are reported by IPEA and sourced from the *Secretaria do Tresouro Nacional* in Brazil. All of these are deflated to 2011 prices. Table 3 presents summary statistics on the FPM transfer for the years 1994-2011, as well as the different levels of spending and revenues for municipalities.

## 3.2 *Censo Escolar*

The *Censo Escolar* is an annual census of schools in Brazil below the tertiary level. The survey, carried out in May, covers both private and public schools, and has been running continuously since 1995. The data are provided by the *Instituto Nacional de Estudos e Pesquisas Educacionais* (INEP). The *Censo Escolar* survey form varies slightly from year to year; however, the general topics which remain consistent over time include basic information, physical and instructional features of the school, teachers and staff, numbers of classes and students, and student flows.

For the purposes of this paper, we wish to measure investments in education. We do this by looking at school resources in the form of teaching staff and classrooms. We consider two measures of teaching staff: the total number of teachers, and teachers' average qualification level. Classrooms are counted twice in the data: available classrooms, and classrooms in use. We retain both of these variables, and create a third to capture the difference between the two, in percent of all available rooms. This final variable gives the excess capacity of classrooms in the municipality, and can shed some light on whether new classrooms are being put to use. These variables are detailed in Appendix A.1.

To give a sense of magnitude, the *Censo Escolar* surveyed 264,735 schools in 2001. To aggregate these school-level variables at the municipal level, we sum the relevant counts over all schools, both public and private. Descriptive statistics for these municipal sums can be found in Table 6. While an analysis using these municipality-wide data can give us a picture of how changes in the federal transfer affect the overall stock of education resources, it may dissimulate heterogenous effects of the transfer across schools of different types. In particular, two types of school heterogeneity are of interest: what grades are offered at the school, and what unit of administration is responsible for the school.

In the early years of our panel, not all municipalities offered classes at all levels of basic education. Table 7 gives the percentage of municipalities which have at least one school offering each level of basic education, for the years 1996, 2001 and 2006. While the availability of each level of instruction is nearly universal by 2006, in 1996 a full 10% of municipalities do not have a single school offering secondary school grades. Children in those municipalities who wished to study beyond eighth grade would have to move away from home in order to do so. In the same year, 0.7% of municipalities did not even have a school offering senior primary school (grades 5-8): while this share is small, it nevertheless represents more than 30 municipalities. As the completion of primary education became more and more common over the last two decades, capacity constraints in secondary schools – and even in senior primary – have become a pressing concern. Indeed, while enrollments in secondary school rose by 1.8 million from 1980 to 1994, they grew by more than twice that from 1994 to 2003 (JBIC, 2005 [22]). It is therefore of interest to measure the expansion of education resources at different levels of education: for a fixed level of resources, shifting teachers and classrooms towards higher levels of instruction can represent a substantial improvement in local education provision.

Of the schools surveyed in 2001, 186,972 were under municipal jurisdiction, 39,329 under state jurisdiction, 38,196 were private schools and 238 were federally-run. Examining the heterogeneous impact of changes in municipal transfers over these different sectors is of interest for two reasons. First, it allows us to explore how changes in transfers affect the private education sector. Second, it allows us to highlight municipality-run schools: since these schools are financed out of the municipal budget, we should expect the greatest impacts of the transfer on these institutions.

We therefore replicate our baseline estimates separately for schools with the characteristics described above.[9] While we will analyse the two types of heterogeneity separately, it is important to keep in mind that they are not unrelated. Since the 1996 education reform law, state school boards were responsible for secondary education, while municipalities were to govern pre-school and primary education. As can be seen in Table 8, this was still far from reality halfway through our panel in 2001. Nevertheless, it must be kept in mind that investments in higher grade levels will be largely under the purview of the state, while investments in younger grades remain (or shift towards) municipal control.

# 4 Estimation Strategy

## 4.1 The Municipal Funding Formula

The *Fundo de Participacao dos Municipios* (FPM), which began in 1984 and was confirmed in the 1988 constitution, is a transfer from the federal government to municipal governments which represents on average around 30% - 40% of municipal revenues each year. The federal government collects both income taxes and industrial product taxes, and 22.5% of these revenues

[9] This process is straightforward for administrative jurisdictions, as each school is run by a single administration (we exclude federal schools, since they are few and quite exceptional). Given that schools may offer classes across more than one of the categories identified in Table 7, we use a hierarchical classification system to map schools to the level of schooling they offer. This is described in Appendix A.2.

are transferred to through the FPM. While 12% of these funds are transferred to state capitals and large municipalities (those with a population greater that 400,000), and a further 6% is transferred to municipalities with a population larger that 156,216, the rest is transferred to the reamaining municipalities according to a strict formula. Throughout our analysis, we focus on these smaller municipalities. The formula, which takes into account municipal population as well as the popultaion of other municipalities within the same state, is defined as follows:

$$FPM_t^{is} = \frac{\theta(pop_{it-1})}{\sum\limits_{j\epsilon s}\theta(pop_{jt-1})} \times S_t, \tag{1}$$

where:

- $FPM_t^{is}$ is the transfer amount received in municipality $i$ which belongs to state $s$ in year $t$.
- $\theta(pop_{it-1})$ is a discrete coefficient determined according to the municipal population in year $t-1$ (see Table 2).
- $\sum\limits_{j\epsilon s}\theta(pop_{jt-1})$ is the sum of the discrete coefficients for all municipalities in the state $s$ in year $t-1$.
- $S_t$ is the total transfer allocated to state $s$ in year $t$, which is determined as a share of total federal revenues from IR and IPI tax incomes, according to an allocation formula based on state populations in 1990.

The FPM received by a municipality in a given year is therefore both a function of its own population, and the population of other municipalities in the same state.

A number of previous studies have exploited the discontinuity created when population growth pushes a municipality over the threshold between two coefficients.[10] There is reason, however, to question the exogeneity of such changes. First, recent evidence suggests that municipalities can in fact influence their own population estimates (see Lischtig (2012) [24]). Second of all, the population thresholds may influence the probability that a municipality will split in two, thereby creating groups of municipalities at the thresholds necessary to receive a certain level of transfers.

[10]See for example Gadenne (2014) [16], Lischtig (2012) [24], Brollo et al. (2013)[10].

It is also important to note that the value of $S_t$ is a function of federal revenues. While every municipality concerned is small compared to the the nation as a whole, and therefore only contributes a small share of these revenues, general trends in the economy may increase (or decrease) $S_t$, while at the same time affecting municipal revenues and spending. One such example would be a change in the nominal interest rate. This would bias upwards an ordinary least squares estimation of the impact of FPM on local government spending and revenues, as important co-determinants of FPM and local government finance would be omitted.

To overcome these issues, our estimation strategy relies on changes in the FPM at the municipal level which arise from changes in the population coefficients other municipalities within the same state. Over the course of the 1990s and 2000s, municipalities both grew in terms of their population, and in terms of their number: over the course of our panel alone, the number of municipalities increases by 560. By changing the population coefficients of the municipalities in question, these changes affect the share of FPM revenues attributed to other municipalities.

Taking natural logs of Equation 1, we see:

$$\begin{aligned}
lnFPM_t^{is} &= ln\theta(pop_{it-1}) - ln\sum_{j\epsilon s}\theta(pop_{jt-1}) + lnS_t \qquad (2)\\
&= lnFPM_{t-1}^{is} + ln\theta(pop_{it-1}) - ln\theta(pop_{it-2})\\
&\quad + lnS_t - lnS_{t-1} - ln\sum_{j\epsilon s}\theta(pop_{jt-1}) + ln\sum_{j\epsilon s}\theta(pop_{jt-2}) \qquad (3)\\
&= lnFPM_{t-1}^{is} + ln(\frac{pop_{it-1}}{pop_{it-2}}) + ln(\frac{S_t}{S_{t-1}}) - ln(\frac{\sum_{j\epsilon s}\theta(pop_{jt-1})}{\sum_{j\epsilon s}\theta(pop_{jt-2})})
\end{aligned}$$

The last term in the above equation is quasi-exogenous and depends on the changes in all the municipalities in the state over the last two years. It nevertheless includes the population coefficient of the municipality under consideration which, as discussed above, is potentially endogenous. Restricting ourselves only to variation in other municipalities' coefficients distills the exogenous elements of the last term.[11] We therefore instrument the annual change in logged FPM in year $t$ using the change in the logged of the sum of these population coefficients in year

[11] Formally: $ln(\frac{\sum_{j\epsilon s}\theta(pop_{jt-1})}{\sum_{j\neq i\epsilon s}\theta(pop_{jt-2})})$.

$t-1$, for all municipalities $j \neq$i in state $s$.[12]

## 4.2 Estimating Equations

We estimate the impact of a change in FPM, instrumented using the exogenous change in population of *other municipalities'* population, on local government spending and revenues, and in turn on education infrastructure and teaching staff. A first approach would be to estimate a simple fixed-effects model, as follows:

$$lnY_{it} = \beta_0^{FE} + \beta_1^{FE} lnFPM_{it} + \beta_2^{FE} lnY_{it-1} + \beta_3^{FE} X_{it} + \beta_4^{FE} X_i + \beta_5^{FE} X_t + \varepsilon_{it} \text{ [Second Stage]} \quad (4)$$

$$lnFPM_t^{is} = \delta_0^{FE} + \delta_1^{FE} ln \sum_{j \epsilon s} \theta(pop_{jt-1}) + \delta_2^{FE} X_{it} + \delta_3^{FE} X_i + \delta_4^{FE} X_t + v_{it} \text{ [First Stage]} \quad (5)$$

where $Y_{it}$ is the outcome variable of interest, $X_{it}$ are time-varying municipal controls, $X_i$ are municipality fixed-effects, and $X_t$ are year controls. $\delta_1$ is expected to be negative: an increase in the population of other municipalities in the state leads to a decrease in the FPM attributed to the municipality in question.

This approach is problematic, however, as Nickell observed (1981 [29]). Because the fixed-effects transformation de-means the data, the modified error term in any period will be correlated with the lagged dependent variable, included in Equation 4 as a regressor. This will bias estimates derived from the equations above, leading to the eponymous Nickell Bias.[13]

Re-casting the equations in first differences goes some way towards remedying the problem: since the data is transformed using only the previous period, the modified error term in a given period will no longer be correlated with *all* lagged values of the dependent variable. If the data follows an AR(1) process, however, the current period error will still be correlated with the first

[12] Recall that the sharing of the transfer among municipalities is based on coefficients determined by their estimated populations in the previous year. This means that shocks to neighbouring municipalities in year $t$ will affect the transfer share in year $t+1$.

[13] Roodman (2009 [31]), Bond (2002 [8]) and Baum (2006 [5]) provide excellent introductions to this issue, and to the approaches outlined below.

lag of the dependent variable (differenced or not), and the bias remains. One way to correct for this bias is to instrument for the first lag of the dependent variable with deeper lags. If the data truly follow an AR(1) process, these deeper lags will not be correlated with the current (differenced) error term. The equations below outline this approach.

$$\triangle lnY_{it} = \beta_0 + \beta_1 \triangle lnFPM_{it} + \beta_2 \triangle lnY_{it-1} + \beta_3 \triangle X_{it} + \beta_4 X_t + \triangle \varepsilon_{it} \text{ [Second Stage]} \tag{6}$$

$$\triangle lnFPM_t^{is} = \delta_0 + \delta_1 \triangle ln \sum_{j\epsilon s} \theta(pop_{jt-1}) + \delta_2 \triangle X_{it} + \delta_3 X_t + \triangle v_{it} \text{ [First Stage I]} \tag{7}$$

$$\triangle lnY_{it-1} = \gamma_0 + \gamma_1 \triangle lnY_{it-2} + \gamma_2 \triangle X_{it} + \gamma_3 X_t + \triangle \eta_{it} \text{ [First Stage II]} \tag{8}$$

where $Y_{it}$ is the outcome variable of interest, $X_{it}$ are time-varying municipal controls, and $X_t$ are year controls. If a more persistent data process is suspected, the twice-lagged $Y_{it}$ in Equation 8 can be replaced with third or fourth lags, although this evidently comes at a cost in term of panel length.

While this approach should eliminate the Nickel Bias from our second stage estimates, the work of Arellano and Bond (1991 [3]), and Holtz-Eakin, Newey and Rosen (1988 [21]), has lead to the development of more efficient methods based on Generalised Methods of Moments. The Arellano-Bond estimator essentially improves on Equations 6-8 by efficiently including all available lags when constructing the instrument in Equation 8. The Arellano-Bond estimator has gained considerable popularity in recent years and, thanks to work such as Roodman (2009 [31]), has become widely accessible.

We first estimate Equations 6-8 using the Anderson-Hsiao first-difference estimator, instrumenting the change in the dependent variable with its own lagged values. This gives two stage least squares regressions in first differences. For our primary analysis, however, we estimate Equations 6-8 using the Arellano-Bond estimator written by Roodman in Stata, xtabond2 (2009 [31]). Using the built-in command to do so, we check the autocorrelation of our residuals and adjust the lags of the dependent variables used as instruments accordingly. Further details about each set of regressions are given below each table of results.

# 5 Results

## 5.1 Ordinary Least Squares: Impacts of Transfers on Revenues and Spending

Although we expect the cross-sectional relationship between transfers and local revenue and spending to differ from the causal relationship, it is illustrative to begin with a simple OLS estimation. Table 9 presents results of an OLS estimation of the impact of a change in FPM on changes in revenues and expenditures, in first differences. In all estimations, we control for the lagged dependant variable and include time dummies. Column 1 shows the direct positive impact of transfers on current revenues. On average, a 1% increase in FPM is associated with a 0.1% increase in current revenues, and vice versa. However, as we observe in columns 2 and 3, there are no impacts on locally-raised tax revenues. Increases in the FPM do not impact the level of tax revenues raised from property or services taxes. Columns 4-6 show the impact of the FPM on municipal government spending. Current spending increases by 0.05% for every 1% increase in transfers. Capital spending increases more, rising by 0.11%, whereas spending on education rises 0.18%.

## 5.2 First Stage Regressions: the Instrument

Do increases in the population coefficients of other municipalities decrease the transfers received by the excluded municipality? In Equation 7 we would expect the coefficient $\delta_1$ to be negative, as shown by the formula for FPM given in Equation 2 above. Table 10 presents first stage results confirming this relationship.[14] As expected, the impact of the logged ratio in state population coefficient is strongly negative and highly significant. The R-squared of 0.26 suggests that we explain approximately a quarter of the variation in FPM through our instrument.

[14]We make use of the user-written xtivreg2 Stata command in our Anderson-Hsiao estimations; see Schaffer (2005 [33]).

## 5.3 Second Stage: Causal Impacts of Transfers on Revenues and Spending

### 5.3.1 Arellano-Bond: Impacts of Transfers on Revenues and Spending

We now proceed to the full estimation, and look at how the FPM impacts local revenue raising and expenditure. In Table 12 we report results from an Arellano-Bond two-step difference GMM estimation, with deviations in first differences,[15] where the change in the aggregate population coefficient of other municipalities in the state is an instrument for the change in FPM. Lagged differences in both the logged transfers and logged dependent variables are included as GMM instruments to correct for autocorrelation. The number of lags included for each estimation are reported in the table.

Column 1 shows a substantial increase in current revenues due to an increase in transfers, with a 1% rise in transfers leading to a 0.4% rise in revenues. On average, this means that a doubling of FPM transfers, or an increase of 600 000 R$, leads to a 888,000 R$ increase in current revenues – about 25% higher than the average level of transfers received. Some additional funds must be coming from elsewhere. This estimate is four times that found in the simple OLS regression. The source of these additional funds appears in column 3, which shows the impact of the transfer on tax revenues. In contrast to the argument that transfers crowd out local revenue collection, the results show that an increase in transfers leads to a *rise* in local services tax revenues. Due to the constraints placed on the levels of these taxes, this effect is most likely driven, at least in part, through increased enforcement. A positive municipal income shock leads to an increase in efforts to collect revenues locally: a crowding-in effect.

Columns 4-6 look at the impacts of the transfer on municipal government spending. For both current and capital spending increases, we observe a particularly large effect on capital spending: a 1% increase in FPM transfers leads to an increase nearly twice as large on capital spending. This may in part explain our results in columns 1 to 3. Increased revenues from transfers reduce financial constraints for local governments. If this means they can now undertake large-scale capital projects, they may raise additional taxes to help fund them. Current spending increases

[15] Table 40 in the appendix replicates our primary estimation using system GMM, instead of GMM in differences as reported throughout the paper. The results are very similar, supporting our estimates and conclusions, although the effects are all somewhat smaller.

by a smaller share of 0.15% for every 1% rise in transfers. However since current spending is on average five times as large as capital spending, a 0.15% rise in current spending is in absolute size represents on average half of the increase in capital spending observed for the same increase in transfers. A third of the increase in revenues are therefore going to current spending, and two thirds to capital spending. Column 6 shows the impact on educational spending, which we discuss below.[16]

Table 13 repeats the previous estimation, with the addition of a lag in the FPM transfer. This allows us to see whether the estimated impacts persist. Column 1 shows that the positive impact of increased FPM on revenues is reversed by 62% the following year - the increase (or decrease) persists somewhat, but decays quickly as tax revenues partially adjust. Column 3 shows that the increase in service tax holds for the following years, but does not increase or decrease further. Interestingly, a negative impact on property taxes, which was present but not significant in the previous regression, is now more pronounced. This effect also reverses in the second year, leaving no long-term impact. When comparing the sizes of these effects, the increase in service tax in the first year more than offsets the fall in property tax; the overall impact is a 0.06% rise in tax revenues for a 1% rise in transfers.[17]

In columns 4-6 we look at the impacts on spending over two years. The positive current expenditure effect observed in the first period is basically fully reversed by the second period. However, capital spending continues to increase: the increase in transfers has a longer-term impact on capital spending.

#### 5.3.2 Anderson-Hsiao: Impacts of Transfers on Revenues and Spending

While our primary analysis is done using the Arellano-Bond estimator, we first demonstrate that our primary results hold using the Anderson-Hsiao first-different estimator. Table 11 presents results of instrumental variables estimations of the impact of a change in FPM on changes in revenues and expenditures, in first differences. In addition to instrumenting FPM transfers with

[16] As a crude check on our Arellano-Bond specification, we estimate upper and lower bounds on the autoregressive coefficients reported in Table 12. Following Bond (2002 [8]), we do this by estimating our equation by both pooled OLS and fixed-effects. The results are reported and discussed in Appendix B.1. Most of the autoregressive coefficients in the table fall within the bounds.

[17] Taking the average level of services and property tax over all years, we see that the impact on property tax is equal to -0.129*597424=-77067, and on services tax is equal to 0.211*791259=166955, leading to a total rise of 89888R$ on average from a 100% rise in transfers, or equivalently, a 6% rise in the level of taxes raised.

other municipalities' population coefficients, we instrument the lagged dependent variable with its 2nd, 3rd and 4th lags.

The estimated coefficients are broadly similar in size, and identical in sign, to those reported in Table 12. Property tax and capital expenditures are the most notably different. Property tax responds strongly negatively to increases in FPM in the Anderson-Hsiao estimation (decreasing by 0.75% for a 1% increase in FPM), while it the coefficient is just barely negative, and statistically indistinguishable from zero, in the Arellano-Bond estimation. Capital expenditure, which increases substantially in the Arellano-Bond estimation, display a small, statistically insignificant increase in the Anderson-Hsiao estimation.

#### 5.3.3 Arellano-Bond: Further Impacts of Transfers on Government Spending

Table 14 reports Arellano-Bond estimates of the impact of a change in FPM on local municipal spending, broken down into finer categories. Spending in all areas increases, although there is considerable variation in the size of this increase. Education and welfare spending both increase by approximately 0.17% for a 1% increase in FPM. Health and sanitation spending rises by 0.6%, justice spending by over 0.7%, and transport spending, as well as housing and urbanism spending, increase by far larger amounts: 1.35% and 2.17%, respectively. Firstly, these latter two increases are unsurprising considering the response observed above in capital spending. An increased FPM transfer leads to a disproportionate rise in capital spending, and these results show that much of this spending is likely happening on transport and urban projects. The increase in spending on justice is also particularly large; however, justice spending is on average very small (1/200th of the size of education spending, and 1/150th that of health spending), so these increases are small in magnitude. In absolute terms, the impact on housing and urbanism is the largest, at a 48,760 R$ increase for a 1% rise in transfers, compared to 28,278 R$ in health spending, 10,707 R$ in education and just 2,172 R$ in justice spending.[18]

[18]These are calculated by multiplying the average levels of expenditure in each sector across all municipalities included in the dataset and across all years by the percentage increases documented above and in Table 14.

#### 5.3.4 Arellano-Bond: Impacts of Transfers on Revenues and Spending, Interacted with Municipality Wealth

As detailed above, municipalities vary greatly in both size and wealth. In our analysis, we focus only on municipalities with populations below 156,216. However, even within this group, there are large inequalities both across and within regions. The transfers are in place in part to target government spending in poorer regions and ensure development across the whole of Brazil. It is therefore particularly important to observe whether the impact of transfers varies according to the wealth of the recipient municipality.

Tables 15 and 16 report results from estimations of the impact of FPM transfers, and of FPM transfers interacted with logged GDP per capita at the start of our panel (in thousands of R$ per person in 1996), on revenues and spending. While the impact on current revenues and property taxes are largely unaffected by the interaction term, the results show that the increase in ISS tax revenue is driven by wealthier municipalities. The crowding-in effect, by which higher transfers leads to greater sourcing of local funds, is only present in wealthier municipalities. In column 5, we observe that the wealthier municipalities also respond to FPM increases with a greater rise in capital spending.

Columns 4 and 6 show that the effects on current spending, in contrast, are focused on poorer municipalities. An increase in FPM results in current spending rising in the poorest municipalities, but less so in richer municipalities.[19] A similar effect is observed in education spending: richer municipalities respond less to an increase in FPM transfers than do poorer ones. The response declines rapidly with increasing initial GDP per capita: for the richest municipalities in our dataset, these coefficients predict that the net effect of increases in transfers on education spending will be negative. These few very wealthy municipalities on average move public spending away from education when transfers increase.

When we look across all spending categories (Table 16), we observe that similar effects occur in welfare spending as in education, although in this case the net effect never becomes negative. Poorer municipalities' welfare spending nevertheless reacts more strongly to an increase in the transfer. For the health, transport and housing and urbanism sectors, however, richer

[19] As the GDP per capita threshold beyond which the effect is negative is never reached within our dataset, the effect remains positive even for the richest municipalities.

municipalities see larger increases in spending than do poorer municipalities.

These effects suggest that while FPM transfers increase spending across the board, for poorer municipalities the effect on current spending dominates, with particularly large effects in education and welfare spending. For richer municipalities, however, the increased transfer goes disproportionately towards increasing capital spending, with large increases in spending in transport and housing and urbanism.[20]

## 5.4 Second Stage: Causal Impacts of Transfers on Education Resources

### 5.4.1 Arellano-Bond: Impacts of Transfers on Education Resources - Municipality-wide effects

Table 17 presents results from our main estimating equation, with a separate equation for each of the five school resource variables described in Section 3. Increases in the FPM transfer increase both the number of teachers employed in the municipality (column 1) and the total number of available classrooms (column 3). For a 10% increase in the transfer, municipalities increase their teaching staff by 0.3%; or, for a municipality with an average number of teachers, by slightly less than one teacher. The effect on available classrooms is even smaller, corresponding to an increase in 0.2 classrooms in response to a 10% increase in transfers. While the number of classrooms increases slightly, the number of classrooms in use decrease by 0.15, leading to an increase in the excess capacity of classrooms, as can be seen in column 5. Table 17 also shows that these changes are associated with an increase in average teacher education: although quite precisely estimated, the effect is extremely small.[21]

Table 18 replicates the regressions in Table 17, now including an interaction between the current year transfer and municipal GDP per capita in 1996. Most of the effects of the transfer shrink towards zero with this control added, with the exception of the regression on number

[20] As the impacts of transfers may also vary according to the politics of the local government, we replicated these estimations while interacting the level of the FPM transfer with various political variables. First, we included the political leaning of the municipal government, and found no significant effect of this on the outcomes discussed here. Secondly, we interacted the FPM transfer with the percentage majority of the incumbent party in the previous municipal election. A locally stronger political party may react differently to an increase in transfers, as they have different incentives with regard to the next election compared to a vulnerable incumbent. However, we find no significant differences in the way FPM impacts local tax revenues and spending in this case either. These results are available on request.

[21] The coefficient of 0.026 implies that a 10% increase in the transfer would induce a single teacher to increase his education by one level.

of teachers in column 1, which doubles in size. The interaction term in column 1 is negative: although it is small, this suggests that increases in teaching staff in response to transfer shocks are more substantial in poorer municipalities.

How big are these changes with respect to changes in funding? At the mean, a 10% increase in the FPM transfer corresponds to approximately 600 000 R$. While our data do not include figures on teacher wages, Alves and de Rezende Pinto (2011 [2]) report that, in 2009, average monthly teacher salaries were 1683 R$ for junior primary teachers, 1856 R$ for senior primary teachers, and 2218 R$ for secondary school teachers.[22] Over twelve months, an additional teacher could reasonably cost somewhere between 20 000 R$ to 27 000 R$, or 3-4% of a 600 000 R$ increase in transfers. The changes in teacher numbers estimated in Tables 17 and 18 are therefore non-negligible investments, particularly given that education is only part – albeit a significant one – of municipal expenses.[23]

We examine the distribution of these effects on schools using two different lenses: first, by breaking schools down according the level of instruction offered,[24] and second, according to the administrative jurisdiction under which the schools operate. This is done separately for the two groups of indicators we consider: teachers and classrooms.

#### 5.4.2 Arellano-Bond: Impacts of Transfers on Teachers - Channels

Tables 19 and 20 show the effect of transfers on the number of teachers, broken down by level of instruction and administration, respectively. The first two columns of both tables reproduce the results from Tables 17 and 18, for comparison; in all tables, even columns include the interaction between the current transfers and municipal GDP per capita. Table 19 shows that there is much more movement of teachers in response to the policy than the aggregate results suggest. Columns 4 and 6 demonstrate a substantial increase in teachers in schools offering upper-level grades: in primary schools teaching grades 5-8 and in secondary schools, a 10% increase in transfers is

[22] We adjust the values reported in Alves and de Rezende Pinto (2011 [2]), which are from 2009, to 2011 R$ using the IPEA deflator. Since the salary of public school teachers depends on their education level, the observed differences in wages across teachers at different levels of instruction can be due either to differential wages across school levels in private schools, or the differences in the average education attainment of teachers at different levels in public schools.

[23] As described in Footnote 16, we preform a crude check on our Arellano-Bond specification for the estimates reported in Table 17 by calculating upper and lower bounds for the autoregressive coefficients. These coefficients all lie within the expected range: see Appendix B.1.

[24] For details on how this subdivision is defined, see Appendix A.2.

associated with a 1.3-1.5% increase in teaching staff. Column 8 gives a hint as to where these teachers are coming from: teaching staff at primary levels 1-4 falls by 1.6% for a 10% increase in transfers. The number of teachers working in preschools, however, appears unresponsive to transfer changes. This re-allocation of teachers is mitigated in richer municipalities: the interaction term is negative for secondary and senior primary schools, while it is positive for junior primary schools.

Table 20 presents the equivalent results, with schools subdivided by administrative jurisdiction. Public schools at both state and municipal levels display an increase in teaching staff, while private schools lose teachers. The sizes of these effects are attenuated, however, compared to those in Table 19: these jurisdictions will have schools teaching classes at different levels, and will therefore represent weighted averages the columns of the former table. Interestingly, the positive effect of the transfers on teaching staff at state schools is diminished in richer municipalities, while in such municipalities the transfer has a stronger impact on municipal schools.

Tables 21 and 22 report the results for teacher education. In contrast to the results on teacher numbers, the small effect of the transfer on the average education of teachers does not strengthen, but rather disappears upon examination at a finer disaggregation. While a few coefficients are precisely estimated, even these remain small and no strong pattern emerges across either levels of instruction or administrative jurisdictions.

These four tables suggest that, although teaching staff was not altered dramatically in response to an increase in municipal transfers, non-trivial changes in staff allocations did take place. In practice this meant an increase in the number of teachers employed at schools offering upper-level grades, particularly in poorer municipalities. While these increases come at the expense of teachers in schools devoted to junior primary education, this nevertheless suggests an overall improvement in educational possibilities, particularly as poorer municipalities are more likely to be constrained in their offerings of higher grade-levels.

### 5.4.3 Arellano-Bond: Impacts of Transfers on Classrooms - Channels

Tables 23 and 24 show the number of existing classrooms, while Tables 25 and 26 present the number of classrooms in use. Looking first at Table 23, we can see that increases in transfers *decrease* the number of classrooms available in schools offering higher grades and preschools, while increasing them substantially in junior primary schools, in direct contrast to what we had seen for teachers. Many of these effects shrink and lose statistical significance, however, when the interaction with GDP is included. For junior and senior primary schools, the interaction term crowds out the main effect, suggesting that richer municipalities are the only ones responding negatively to transfer increases. Secondary schools, on the other hand, increase their classrooms when the interaction term is added, consistent with the trend in teacher numbers: for a 10% increase in the transfer, secondary schools increase their classroom count by 0.7%. The number of classrooms in preschools retains the negative sign, even when the interaction term is included: the number of preschool classrooms falls by 2% following an increase in transfers of 10% (column 10). Table 24 suggests that state schools increased their number of classrooms, and that this effect is largely uncorrelated with municipal GDP. Municipal schools show a smaller decrease, while private schools in wealthy municipalities grew.

Are these classrooms being put to use? Tables 25 and 26 are reassuringly similar to the previous pair: the number of classrooms in use follows a similar trend to the number of classrooms available. The most striking difference appears in the aggregated regressions: while available classrooms increase slightly, the number of classrooms in use decreases by a similarly small amount. Examining these changes in more detail suggests that this difference in the aggregate is not due to substantially different trends, but rather the aggregation of effects that are on average close to zero.

Finally, Tables 27 and 28 show the difference between the number of available classrooms and the number of classrooms in use, expressed as a percentage of available rooms. While the coefficients across both tables are nearly ubiquitously positive, these increases are very small: increases in transfers increase the share of unused classrooms, but not substantially.

## 6 Discussion

This paper contributes to the literature of public finance and the impact of transfers on local tax revenue raising and spending patterns. We find causal evidence of a Flypaper Effect in Brazilian municipal finance, whereby intergovernmental transfers result in increased public spending at the local level. Furthermore, increased transfers lead to greater local tax revenue collection, displaying substantial crowding-in. These increases in spending translate into small improvements in school resources, as well as the re-allocation of resources towards schools offering higher levels of instruction. In this section we will discuss how these results relate to several recent papers on similar topics.

In their study of municipal finance in Sweden, Dalhberg et al. (2008 [12]) find results consistent with a 100% pass-through rate of federal grants to municipal spending. They find no evidence, however, of changes in taxes in response to changes in transfers. Our results are particularly striking in comparison due to the positive effect of transfers on taxes: this increase in taxes 'crowds-in' the transfer, so that a 1 R$ increase in transfers raises revenues by 1.47 R$. Spending increases as well, with a 1 R$ increase in the transfer leading to a 1.25 R$ increase in spending.[25]

Our finding of a substantial Flypaper Effect in Brazilian municipalities is consistent with Gardner's (2013) [17] study of municipal finance in the Northeastern region. Although the variation in revenue-sharing she exploits is different from our own, she finds no evidence for the crowd-out municipal revenue-raising following increases in federal transfers. Gardner does not find any causal evidence for an increase in taxes due to the transfer, although she does detect a positive relation in the cross-section.

In contrast with our finding of a small but positive increase in education infrastructure and teaching staff, Gadenne (2014) [16] finds no change in classroom number or quality in response to changes in the FPM. In addition to methodological considerations, two differences between our approaches should be noted.[26] First, Gadenne's results are based on classrooms *in use*, rather than the physical stock of available classrooms. While we find an overall negative effect of the transfer on classrooms in use, this effect weakens and becomes statistically insignificant when the interaction with initial municipal GDP per capita is added. Second, while we include all

[25]These back-of-the-envelope calculations are based on mean values from Table 4 and the coefficients reported in Table 12. A 1% increase of transfers at the mean is equal to 60 000 R$. This leads to a 0.4% increase in revenues, which translates into an increase of 88 000 R$ at the mean, or a unit increase of 1:1.47. Capital spending increases by 53 700 R$, while current spending increases by 21 500 R$, for a total spending increase of 75 000 R$, or a spending increase of 1:1.25.

[26]While both papers exploit the FPM rule for identification, Gadenne (2014) uses a regression discontinuity design based on own-municipality population, while we derive an instrument based on the population of other municipalities in the same state. Furthermore, although the variables used overlap, we measure the log of number of classrooms, while Gadenne reports the number of classrooms per school-aged child. Finally, our panels span slightly different time periods (1996-2006, in our case, versus 1998-2009).

schools, Gadenne restricts her analysis to municipality-run schools. While these schools should, in theory, be the most responsive to changes in municipal funding, they are also more likely to be primary schools (see Table 8). Since the introduction of the FUNDEF in 1998, funding to primary schools is largely determined by revenue-sharing rules which equalise funding on a per-capita basis across the state. Although municipalities were free to contribute additional resources on top of this amount, to the extent that the funding rules were binding, increases in the FPM which resulted from changes in the population coefficients should not affect funding to primary schools.[27] In states where the funding rule did not bind we would indeed expect changes in the transfer amount to affect local spending, even at the primary level; however, these could be very small in aggregate.

Nevertheless, even our most comparable estimates do not line up with those of Gadenne. In Table 26, columns 5 and 6 report on results of a regression of logged classrooms in use in municipality-run schools only. Both overall, and when an interaction of the transfer with GDP per capita is included, we find a negative effect of transfers on classrooms in use at municipal schools. Ultimately, this negative effect – and indeed, the small fall in the number of classrooms in use overall – should not overshadow the substantial re-allocation we find across schools. Returning to Table 26, for example, we can see that the fall in municipal-school classrooms is compensated for by an increase in state-school classroom. More importantly, it is also associated with a much larger increase in the number of classroom in use in secondary schools (see Table 25).

We cannot comment extensively on the relation of our results to those of Bastos & Straume (2013) [4]. While their finding that increases in municipal transfers cause an expansion of municipality-run preschools appears to contrast with our own results – that is, that preschool classrooms decrease when transfers increase – a direct comparison is not warranted. Our hierarchical classification of schools according to the levels of instruction offered is designed primarily to capture expansions into upper levels of instruction. It therefore classifies a significant fraction

[27] To understand why this is the case, consider the funding thresholds imposed by the FUNDEF. In essence, the FUNDEF stipulated that 15% of the primary municipal transfers (including the FPM) must be contributed to a state-wide fund. This fund was then redistributed to public school districts based on the number of students enrolled in primary schools. Exogenous variation in a municipality's FPM transfer comes from changes in the population thresholds (either own population, as in Gadenne (2014) [16] and Litschig & Morrison (2013) [25], or other municipalities' populations, as in our own analysis). Since this does not affect the overall FPM but merely its distribution within the state, it should have no impact on the funds collected by the state FUNDEF. Exogenous changes in FPM should therefore not affect the resources a municipality receives from the FUNDEF: resources which are destined toward primary schools.

of preschools as senior primary schools (all schools which offer grades 5-8, but do not offer secondary school, are classified as senior primary schools; see Appendix A.2). For this reason, we do not give much weight to our findings with respect to preschools.

## 7 Conclusion

Despite the host of funding rules designed to improve education in Brazil, serious concerns have been raised over whether public funds are reaching their intended destination. As a 2005 study of the education sector notes:

> *The [1988] Constitution mandated that 25% of state and municipal income and 18% of federal government income go to education. In the years that followed, non-commitment to these constitutional provisions became generalized. The law had reserved the resources, but had not introduced efficient supervisory and control mechanisms. State and municipal governments made use of all types of artifices to include other administrative expenses in the educational budget. The accounting laws made it possible to conceal these artifices, through which resources that should have been invested in education disappeared.* (JBIC et al., 2005 [22] p. 10)

While these concerns are not unfounded, our results are supportive of a non-trivial Flypaper Effect: transfers do increase local spending and are not entirely passed on to the tax payer through reduced taxes.[28] These spending increases are associated with small net investments in education infrastructure and staff, and with more substantial re-allocation of resources across levels of instruction.

We demonstrate that transfers to municipalities in Brazil do in fact 'stick' by first observing that current spending increases contemporaneously with increases in federal funding. While property tax revenues fall, this effect is more than compensated by a rise in revenues from service taxes, with the latter enduring in the subsequent year. This last result is suggestive of a crowding-in effect with respect to service taxes, whereby increased transfers lead to higher tax collection efforts.

[28] Bird and Smart (2002 [6]) discuss issues surrounding transfers in developing countries. For an extended treatment of the Flypaper Effect, see Hines and Thaler (1995)[20].

We are able to take this observation one step further, however, and identify a causal relationship between these changes thanks to exogenous, formula-based, variation in federal transfers. Our analysis reveals a complex pattern of changes in local spending. First, increases in transfers lead to a disproportionate rise in capital spending in the wealthier municipalities. This spending seems to target particularly the transport sector, as well as housing and urbanism, and is financed by both the transfer and an increase in tax revenues. The mechanism driving this increase is not obvious: are local governments able to invest in more costly capital projects now that their budget constraints have been softened? Such a scenario could explain why they seek additional financial resources in response to an increase in transfers. This is an area which has potential for future exploration, particularly if data on actual investment projects at the local level can be assembled.

In contrast, poorer municipalities respond to higher transfers by raising spending on education and welfare. Such an effect does not appear to be politically determined, but more directly a function of the historic average income of local residents. Poorer municipalities are likely to have lower initial educational levels, worse job opportunities, and greater unemployment and health issues. It is therefore unsurprising that increased finances may be directed into funding these areas.

The crowding-in effect mentioned above is present across municipalities irrespective of their wealth. Richer municipalities increase their capital spending following an increase in transfers, financed by both the transfer *and* increased tax revenues. Poorer municipalities increase their current spending, although capital spending is also affected. While there is some evidence for a decrease in property taxes following a rise in transfers, this fall is temporary; the offsetting increase in service taxes, however, endures. These results suggest that intergovernmental transfers to local governments do indeed stick, and that the increased revenues are largely spent immediately. In poorer municipalities these effects are particularly felt in education and welfare, whereas wealthier municipalities spend more on capital projects.

Is this increased spending having any effect? Our results show small net increases in the number of classrooms and in teaching staff, along with substantial movement of resources across schools offering different levels of instruction. Teacher numbers grow in schools offering secondary school classes and senior primary grades 5-8, but shrink in schools offering only junior primary grades 1-4. These effects are stronger in poorer municipalities. Classrooms, both the total available and the number in use, also increase in secondary schools; however, at least in wealthier municipalities, they decrease in senior primary schools. Junior primary schools, which should be quite well developed, show a substantial increase, particularly in wealthier municipalities. To the extent that higher grade levels remain under-supplied in Brazil, these changes represent genuine improvements in the education system. How these adjustments are taking place, in particular when the redistribution of resources takes place across administrative jurisdictions, is an important question for further research.

## References


[1] J R Afonso and L De Mello. Brazil: an evolving federation. *Managing fiscal decentralization*, 2000.

[2] Thiago Alves and José Marcelino de Rezende Pinto. Teachers' Salaries and Labour Conditions in Brazil: a Contribution From Censuses Data. *Cadernos de Pesquisa*, 41(143):606–639, 2011.

[3] Manuel Arellano and Stephen Bond. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297, April 1991.

[4] Paulo Bastos and Odd Rune Straume. Preschool education in Brazil: Does public supply crowd out private enrollment? *Mimeo*, pages 1–29, 2013.

[5] Christopher F Baum. *An Introduction to Modern Econometrics Using Stata*. Stata Press, August 2006.

[6] R M Bird and M Smart. Intergovernmental fiscal transfers: International lessons for developing countries. *World Development*, 30(6):899–912, 2002.

[7] Olivier J Blanchard, Florencio López-de Silanes, and Andrei Shleifer. What Do Firms Do with Cash Windfalls? *Journal of Financial Economics*, 36:337–360, 1994.

[8] Stephen R Bond. Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, 1(2):141–162, August 2002.

[9] David F Bradford and Wallace E Oates. Towards a Predictive Theory of Intergovernmental Grants. *The American Economic Review*, 61(2):440–448, May 1971.

[10] Fernanda Brollo, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini. The Political Resource Curse. *American Economic Review*, 103(5):1759–1796, August 2013.

[11] B. Bruns, D. Evans, and J. Luque. Achieving world-class education in Brazil: the next agenda. Technical report, 2010.

[12] Matz Dahlberg, Eva Mörk, Jørn Rattsø, and Hanna Ågren. Using a discontinuous grant rule to identify the effect of grants on local taxes and spending. *Journal of Public Economics*, 92(12):2320–2335, December 2008.

[13] Maria Helena Guimarães de Castro. Educational Content and Learning Strategies for Living Together in the 21st Century. In *Forty-sixth Session of the International Conference on Education*, pages 1–138, Geneva, September 2001.

[14] William Easterly. *The White Man's Burden*. Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good. Penguin, March 2006.

[15] Fernanda Estevan. Public Education Expenditures and Private School Enrollment. *Working paper*, pages 1–24, September 2014.

[16] Lucie Gadenne. Tax Me, But Spend Wisely: Sources of Public Finance and Government Accountability. *Mimeo*, pages 1–51, August 2014.

[17] Rachel Elizabeth Gardner. *Essays on Municipal Public Finance in Brazil*. PhD thesis, May 2013.

[18] Nora Gordon. Do federal grants boost school spending? Evidence from Title I. *Journal of Public Economics*, 88(9-10):1771–1792, August 2004.

[19] Nora Gordon and Emiliana Vegas. Education Finance Equalization, Spending, Teacher Quality and Student Outcomes: The Case of Brazil's FUNDEF. In Emiliana Vegas, editor, *Incentives to Improve Teaching Lessons from Latin America*, pages 151–186. Mimeo, Washington, DC, 2005.

[20] James R Hines Jr. and Richard H Thaler. Anomalies: The Flypaper Effect. *The Journal of Economic Perspectives*, 9(4):217–226, October 1995.

[21] Douglas Holtz-Eakin, Whitney Newey, and Harvey S Rosen. Estimating Vector Autoregressions with Panel Data. *Econometrica*, 56(6):1371–1395, November 1988.

[22] Japan Bank for International Cooperation, Paulo Renato Souza Consultores, Tendências Consultoria Integrada, and Núcleo de Estudos de Políticas Públicas da Universidade Es-

tadual de Campinas. Sector Study for Education in Brazil. Technical report, JBIC Sector Study Series, November 2005.

[23] Rosalind Levačić and Peter Downes. Formula funding of schools, decentralization and corruption: a comparative analysis. *International Institutde for Educational Planning*, pages 1–226, 2004.

[24] Stephan Litschig. Are rules-based government programs shielded from special-interest politics? Evidence from revenue-sharing transfers in Brazil. *Journal of Public Economics*, 96(11):1047–1060, 2012.

[25] Stephan Litschig and Kevin M Morrison. The Impact of Intergovernmental Transfers on Education Outcomes and Poverty Reduction. *Mimeo*, pages 1–49, February 2013.

[26] Ricardo A Madeira. The Effects of Decentralization on Schooling: Evidence From the Sao Paulo State's Education Reform. *Working paper*, pages 1–48, October 2012.

[27] Naercio Menezes-Filho and Elaine Pazello. Do teachers' wages matter for proficiency? Evidence from a funding reform in Brazil. *Economics of Education Review*, 26(6):660–672, December 2007.

[28] Dambisa Moyo. *Dead Aid*. Why aid is not working and how there is another way for Africa. Penguin UK, August 2011.

[29] S Nickell. Biases in Dynamic-Models with Fixed Effects. *Econometrica: Journal of the Econometric Society*, 49(6):1417–1426, 1981.

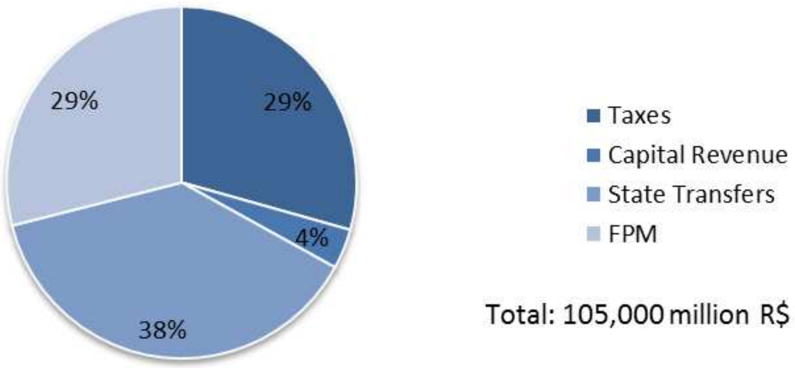
[30] OECD. BRAZIL – Country Note –Results from PISA 2012. Technical report, November 2013.

[31] David Roodman. How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, 9(1):86–136, 2009.

[32] Jeffrey Sachs. *The End of Poverty*. How We Can Make it Happen in Our Lifetime. Penguin UK, November 2011.

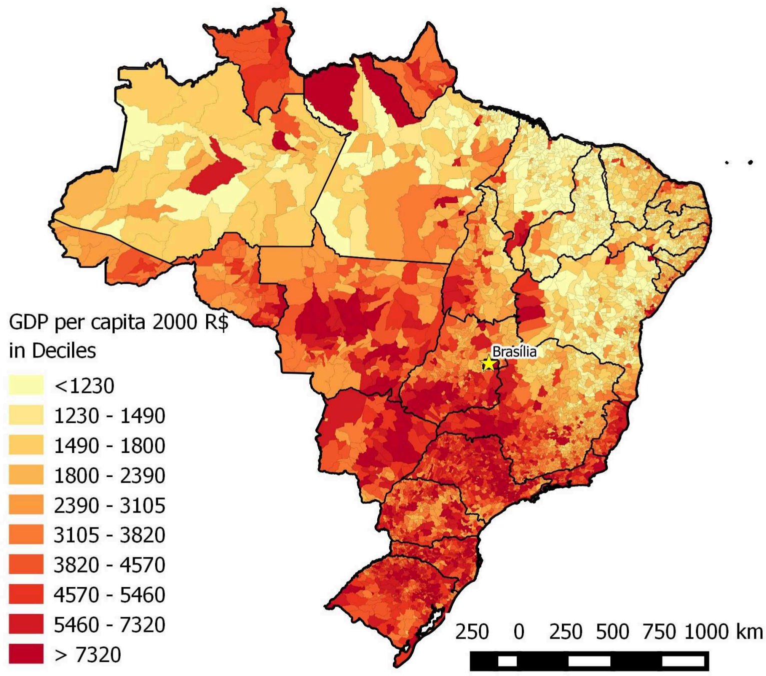
[33] Mark E Schaffer. XTIVREG2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. Technical report, November 2005.

# Tables and Figures




Municipal Revenues by Source, 2005, including all municipalities.

Figure 1: Municipal Revenues by Source (2005)




Map of municipalities in Brazil in 2000, showing GDP per capita in R$ grouped into deciles. State boundaries are also marked.

Figure 2: Municipalities by GDP per capita

Table 1: **Main Taxes by Level of Government**

| **Municipal** | |
|---|---|
| IPTU | Tax on property of urban buildings and land. |
| ISS | On services (municipal, excl. ICMS). Hotels, doctors, schools, entertainment. |
| ITBI | Tax on transfers of real estate. |
| IRFF | Imported Services. |
| **State** | |
| ICMS | State tax on services and goods. Value-Added tax. |
| ITCD | Inheritance. |
| IPVA | Tax on vehicle owners. |
| **Federal** | |
| ITR | Rural Land Tax (minimal revenues). |
| IR | National Income. |
| II / IE | Imports/Exports. |
| IPI | Industrialized products (origin based). |
| IOF | Credit and Insurance. Other Large fortunes (not yet implemented). |
| Social Contributions | not subject to redistribution requirements |

Table 2: **Population Coefficients**

| **Population from** | **to** | **Coefficient** |
|---|---|---|
| 0 | 10188 | 0.6 |
| 10189 | 13584 | 0.8 |
| 13585 | 16980 | 1 |
| 16981 | 23772 | 1.2 |
| 23773 | 30564 | 1.4 |
| 30565 | 37356 | 1.6 |
| 37357 | 44148 | 1.8 |
| 44149 | 50940 | 2 |
| 50941 | 61128 | 2.2 |
| 61129 | 71316 | 2.4 |
| 71317 | 81504 | 2.6 |
| 81505 | 91692 | 2.8 |
| 91693 | 101880 | 3 |
| 101881 | 115464 | 3.2 |
| 115465 | 129048 | 3.4 |
| 129049 | 142632 | 3.6 |
| 142633 | 156216 | 3.8 |

Table 3: **Summary Statistics - FPM**

| Year | Mean FPM (2011 R$) | SD | N |
|---|---|---|---|
| 1994 | 4387354 | 2634123 | 4798 |
| 1995 | 4552415 | 2690028 | 4835 |
| 1996 | 4134047 | 2527073 | 4830 |
| 1997 | 4557121 | 2707523 | 4827 |
| 1998 | 3679038 | 2455156 | 5367 |
| 1999 | 4036580 | 2711055 | 5367 |
| 2000 | 5054309 | 3302217 | 5362 |
| 2001 | 5571110 | 3358467 | 5361 |
| 2002 | 6045659 | 3694150 | 5410 |
| 2003 | 5664848 | 3474111 | 5406 |
| 2004 | 5765282 | 3542602 | 5402 |
| 2005 | 6509011 | 4034838 | 5401 |
| 2006 | 7073519 | 4370917 | 5400 |
| 2007 | 7770415 | 4830194 | 5400 |
| 2008 | 8391647 | 5209521 | 5400 |
| 2009 | 8090079 | 4956257 | 5396 |
| 2010 | 7774450 | 4812532 | 5394 |
| 2011 | 8294548 | 5236320 | 5394 |
| Overall | 6003552 | 4148998 | 94750 |

Table 4: **Summary Statistics - Revenues and Expenditure**

| | (1)<br>Current Revenue | (2)<br>Property Tax Revenue | (3)<br>ISS Tax Revenue | (4)<br>Current Expenditure | (5)<br>Capital Expenditure | (6)<br>Education Spending |
|---|---|---|---|---|---|---|
| mean | 2.22e+07 | 597424.1 | 791259.1 | 1.43e+07 | 2866660 | 6261155 |
| sd | 4.08e+07 | 2799590 | 4048795 | 2.89e+07 | 6858461 | 1.07e+07 |
| N | 84489 | 84489 | 84489 | 89290 | 89290 | 79216 |

Values all in R$ deflated to 2011 levels. All years combined.

Table 5: **Summary Statistics - Municipal Spending**

| | (1)<br>Welfare spending | (2)<br>Education Spending | (3)<br>Justice spending | (4)<br>Housing and Urbanism spending | (5)<br>Health and Sanitation spending | (6)<br>Transport spending |
|---|---|---|---|---|---|---|
| mean | 1255986 | 6261155 | 35709.25 | 2248955 | 4697284 | 963844.5 |
| sd | 3330500 | 1.07e+07 | 1003047 | 5450329 | 8788692 | 2251244 |
| N | 79216 | 79216 | 46217 | 79216 | 83872 | 83872 |

Values all in R$ deflated to 2011 levels. All years combined.

Table 6: **Summary Statistics: Educational Inputs**

| **Variable** | **Mean** | **Std. Dev.** | **N** |
|---|---|---|---|
| Number of teachers | 249.299 | 300.299 | 53898 |
| Teacher Qualification | 2.356 | 0.399 | 53370 |
| Classrooms (available) | 128.593 | 142.101 | 58742 |
| Classrooms (in use) | 121.449 | 134.199 | 58743 |
| Percent excess classrooms | 0.058 | 0.064 | 58737 |

*Source: Censo Escolar* 1996-2006.

Table 7: **Availability of Education Levels Within Municipalities**

| **Year** | **Preschool** | **Jr Primary** | **Sr Primary** | **Secondary** |
|---|---|---|---|---|
| 1996 | 0.966 | 1.000 | 0.993 | 0.901 |
| 2001 | 0.996 | 1.000 | 0.999 | 0.953 |
| 2006 | 0.998 | 1.000 | 1.000 | 0.989 |

*Source: Censo Escolar* 1996-2006. Preschool includes creche and classes prior to first grade; junior primary includes grades 1-4; senior primary includes grades 5-8; secondary school last three years (in general), and includes both standard and professional courses.

Table 8: **Levels of Instruction Offered by School Jurisdiction (2001)**

| | **Preschool** | **Jr Primary** | **Sr Primary** | **Secondary** |
|---|---|---|---|---|
| State | 5796 | 24741 | 21763 | 12815 |
| Federal | 28 | 26 | 36 | 162 |
| Municipal | 75656 | 124601 | 19294 | 947 |
| Private | 27412 | 17841 | 9998 | 6305 |
| Total | 108892 | 167209 | 51091 | 20229 |

*Source: Censo Escolar* 2001. Preschool includes creche and classes prior to first grade; junior primary includes grades 1-4; senior primary includes grades 5-8; secondary school last three years (in general), and includes both standard and professional courses.

Table 9: **OLS - First Differences**

| | (1)<br>D.Log Current Revenue | (2)<br>D.Log Property Tax Revenue | (3)<br>D.Log ISS Tax Revenue | (4)<br>D.Log Current Expenditure | (5)<br>D.Log Capital Expenditure | (6)<br>D.Log Education Spending |
|---|---|---|---|---|---|---|
| D.Log FPM transfer | 0.109***<br>(0.0109) | -0.0293<br>(0.0349) | -0.0557<br>(0.0340) | 0.0491***<br>(0.0105) | 0.116***<br>(0.0245) | 0.176***<br>(0.0158) |
| LD.Log Current Revenue | -0.211***<br>(0.0294) | | | | | |
| LD.Log Property Tax Revenue | | -0.370***<br>(0.00729) | | | | |
| LD.Log ISS Tax Revenue | | | -0.259***<br>(0.00702) | | | |
| LD.Log Current Expenditure | | | | -0.280***<br>(0.0209) | | |
| LD.Log Capital Expenditure | | | | | -0.426***<br>(0.00465) | |
| LD.Log Education Spending | | | | | | -0.323***<br>(0.0169) |
| Constant | 0.0687***<br>(0.00170) | 0.0522***<br>(0.00449) | 0.155***<br>(0.00356) | 0.0479***<br>(0.00115) | 0.0897***<br>(0.00329) | 0.0790***<br>(0.00177) |
| Observations | 67921 | 62728 | 66372 | 72055 | 57546 | 54834 |

Reports OLS estimation of log revenues and expenditures in municipality i and state s at time t, estimated as a function of the instrument (log of state population coefficient). Estimation in first differences with year dummies included, as well as the lagged dependent variable. Standard errors clustered at the municipal level are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Table 10: **First Stage**

| | (1)<br>D.Log FPM transfer |
|---|---|
| LD.lnkothers | -1.698*** |
| | (0.0567) |
| Constant | 0.0546*** |
| | (0.000489) |
| Observations | 84509 |
| Adjusted $R^2$ | 0.263 |

Reports First Stage estimation of log FPM in Municipality i and State s at time t, estimated as a function of the instrument Log of State Population Coefficient. Year dummy controls are included. Anderson-Hsiao estimator in first differences using xtivreg2 in Stata. Standard errors clustered at the municipal level are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Table 11: **Two Stage Least Squares - First Differences**

| | (1) D.Log Current Revenue | (2) D.Log Property Tax Revenue | (3) D.Log ISS Tax Revenue | (4) D.Log Current Expenditure | (5) D.Log Capital Expenditure | (6) D.Log Education Spending |
|---|---|---|---|---|---|---|
| D.Log FPM transfer | 0.611*** | -0.745*** | 0.475** | 0.107* | 0.349 | 0.445*** |
| | (0.183) | (0.212) | (0.228) | (0.0574) | (0.292) | (0.107) |
| LD.Log Current Revenue | 0.409* | | | | | |
| | (0.235) | | | | | |
| LD.Log Property Tax Revenue | | 0.120*** | | | | |
| | | (0.0254) | | | | |
| LD.Log ISS Tax Revenue | | | 0.297*** | | | |
| | | | (0.0299) | | | |
| LD.Log Current Expenditure | | | | 0.237** | | |
| | | | | (0.117) | | |
| LD.Log Capital Expenditure | | | | | -0.0157 | |
| | | | | | (0.0157) | |
| LD.Log Education Spending | | | | | | 0.0346*** |
| | | | | | | (0.0124) |
| Constant | 0.0235 | 0.0474*** | 0.0730*** | 0.0337*** | 0.156*** | 0.0551*** |
| | (0.0164) | (0.00539) | (0.00715) | (0.00571) | (0.00727) | (0.00145) |
| Observations | 49210 | 44518 | 47777 | 52465 | 26695 | 36842 |
| Kleibergen-Paap rk Wald F statistic | 256.7 | 276.9 | 249.3 | 280.7 | 108.8 | 178.5 |

Two Stage Least Squares estimation of log revenues and expenditures in municipality i and state s at time t, estimated as a function of the log transfer to the municipality. The transfer is instrument by lagged log of the state population coefficient. Equations are estimated using the Anderson-Hsiao estimator in first differences via xtivreg2, including year dummies as well as the lagged revenue or expenditure. Standard errors clustered at the municipal level are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Table 12: **Arellano-Bond Estimation**

| | (1)<br>Log Current Revenue | (2)<br>Log Property Tax Revenue | (3)<br>Log ISS Tax Revenue | (4)<br>Log Current Expenditure | (5)<br>Log Capital Expenditure | (6)<br>Log Education Spending |
|---|---|---|---|---|---|---|
| Log FPM transfer | 0.411*** | -0.0514 | 0.257*** | 0.149*** | 1.874*** | 0.171*** |
| | (0.0212) | (0.0590) | (0.0693) | (0.0116) | (0.0771) | (0.0336) |
| L.Log Current Revenue | 0.553*** | | | | | |
| | (0.00966) | | | | | |
| L.Log Property Tax Revenue | | 0.310*** | | | | |
| | | (0.0595) | | | | |
| L.Log ISS Tax Revenue | | | 0.774*** | | | |
| | | | (0.0180) | | | |
| L.Log Current Expenditure | | | | 0.788*** | | |
| | | | | (0.0111) | | |
| L.Log Capital Expenditure | | | | | -0.503*** | |
| | | | | | (0.0211) | |
| L.Log Education Spending | | | | | | 0.619*** |
| | | | | | | (0.0144) |
| Observations | 67921 | 62728 | 66372 | 72055 | 57546 | 54834 |
| Lags | 4-8 | 3-7 | 3-7 | 5-9 | 5-9 | 5-9 |

Arellano-Bond estimation of log revenues and expenditures in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. IV-style instruments are the lagged log of the state population coefficient and year dummies, while lagged values of the FPM transfer and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 13: **Arellano-Bond Estimation with Lag**

| | (1) Log Current Revenue | (2) Log Property Tax Revenue | (3) Log ISS Tax Revenue | (4) Log Current Expenditure | (5) Log Capital Expenditure | (6) Log Education Spending |
|---|---|---|---|---|---|---|
| Log FPM transfer | 0.468 *** | -0.129 ** | 0.211 *** | 0.191 *** | 2.109 *** | 0.200 *** |
| | (0.0220) | (0.0595) | (0.0698) | (0.0209) | (0.0805) | (0.0354) |
| L.Log FPM transfer | -0.292 *** | 0.177 *** | 0.0542 | -0.210 *** | 0.457 *** | -0.0368 |
| | (0.0217) | (0.0399) | (0.0512) | (0.0192) | (0.0625) | (0.0322) |
| L.Log Current Revenue | 0.749 *** | | | | | |
| | (0.0176) | | | | | |
| L.Log Property Tax revenue | | 0.344 *** | | | | |
| | | (0.0587) | | | | |
| L.Log ISS Tax revenue | | | 0.790 *** | | | |
| | | | (0.0196) | | | |
| L.Log Current Expenditure | | | | 0.913 *** | | |
| | | | | (0.0170) | | |
| L.Log Capital Expenditure | | | | | -0.571 *** | |
| | | | | | (0.0223) | |
| L.Log Education Spending | | | | | | 0.647 *** |
| | | | | | | (0.0284) |
| Observations | 67550 | 62530 | 66039 | 67493 | 52994 | 54470 |
| Lags | 4-8 | 3-7 | 3-7 | 5-9 | 5-9 | 5-9 |

Arellano-Bond estimation of log revenues and expenditures in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. IV-style instruments are the lagged log of the state population coefficient and year dummies, while lagged values of the FPM transfer and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 14: **Arellano-Bond - Municipal Spending**

| | (1) Log Welfare Spending | (2) Log Education Spending | (3) Log Justice Spending | (4) Log Housing and Urbanism Spending | (5) Log Health and Sanitation Spending | (6) Log Transport Spending |
|---|---|---|---|---|---|---|
| Log FPM transfer | 0.173 *** | 0.171 *** | 0.793 ** | 2.168 *** | 0.603 *** | 1.355 *** |
| | (0.0573) | (0.0337) | (0.345) | (0.0863) | (0.0442) | (0.0841) |
| L.Log Welfare Spending | 0.627 *** | | | | | |
| | (0.0263) | | | | | |
| L.Log Education Spending | | 0.658 *** | | | | |
| | | (0.0144) | | | | |
| L.Log Justice Spending | | | -0.271 *** | | | |
| | | | (0.0426) | | | |
| L.Log Housing and Urbanism Spending | | | | -0.493 *** | | |
| | | | | (0.0451) | | |
| L.Log Health and Sanitation Spending | | | | | 0.686 *** | |
| | | | | | (0.0195) | |
| L.Log Transport Spending | | | | | | -0.688 *** |
| | | | | | | (0.0449) |
| Observations | 53799 | 54834 | 2033 | 52425 | 59076 | 47847 |
| Lags | 3-7 | 5-9 | 1-5 | 6-10 | 6-10 | 6-10 |

Arellano-Bond estimation of log expenditures in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. IV-style instruments are the lagged log of the state population coefficient and year dummies, while lagged values of the FPM transfer and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 15: **Arellano-Bond with GDP/Capita Interaction**

| | (1)<br>Log Current Revenue | (2)<br>Log Property Tax Revenue | (3)<br>Log ISS Tax Revenue | (4)<br>Log Current Expenditure | (5)<br>Log Capital Expenditure | (6)<br>Log Education Spending |
|---|---|---|---|---|---|---|
| Log FPM transfer | 0.125** | -0.0455 | 0.0501 | 0.187*** | 0.462*** | 0.274*** |
| | (0.0582) | (0.0939) | (0.0775) | (0.0191) | (0.0931) | (0.0353) |
| Ln FPM * Ln GDP/cap in 1996 | -0.00899 | 0.0566 | 0.0544** | -0.0315*** | 0.265*** | -0.0804*** |
| | (0.00685) | (0.0360) | (0.0234) | (0.00654) | (0.0446) | (0.0100) |
| L.Log Current Revenue | 0.537*** | | | | | |
| | (0.0273) | | | | | |
| L.Log Property Tax Revenue | | 0.270*** | | | | |
| | | (0.0561) | | | | |
| L.Log ISS Tax Revenue | | | 0.792*** | | | |
| | | | (0.0175) | | | |
| L.Log Current Expenditure | | | | 0.811*** | | |
| | | | | (0.0115) | | |
| L.Log Capital Expenditure | | | | | -0.273*** | |
| | | | | | (0.0230) | |
| L.Log Education Spending | | | | | | 0.575*** |
| | | | | | | (0.0153) |
| Observations | 61928 | 57899 | 60564 | 66065 | 53092 | 49396 |
| Lags | 2-6 | 3-7 | 3-7 | 3-7 | 4-8 | 5-9 |

Arellano-Bond estimation of log revenues and expenditures in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The revenues and expenditures are estimated as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996. IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996, and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996, and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 16: **Arellano-Bond with GDP/Capita Interaction - Municipal Spending**

| | (1) Log Welfare Spending | (2) Log Education Spending | (3) Log Justice Spending | (4) Log Housing and Urbanism Spending | (5) Log Health and Sanitation Spending | (6) Log Transport Spending |
|---|---|---|---|---|---|---|
| Log FPM transfer | 0.275 *** | 0.274 *** | 0.397 | 0.823 *** | 0.211 *** | 0.659 *** |
| | (0.0617) | (0.0355) | (0.835) | (0.103) | (0.0414) | (0.139) |
| Ln FPM * Ln GDP/cap in 1996 | -0.0484 ** | -0.0804 *** | 0.0224 | 0.261 *** | 0.0305 ** | 0.107 * |
| | (0.0221) | (0.00966) | (0.322) | (0.0415) | (0.0128) | (0.0552) |
| L.Log Welfare Spending | 0.572 *** | | | | | |
| | (0.0293) | | | | | |
| L.Log Education Spending | | 0.601 *** | | | | |
| | | (0.0149) | | | | |
| L.Log Justice Spending | | | -0.153 | | | |
| | | | (0.150) | | | |
| L.Log Housing and Urbanism Spending | | | | -0.304 *** | | |
| | | | | (0.0482) | | |
| L.Log Health and Sanitation Spending | | | | | 0.651 *** | |
| | | | | | (0.0207) | |
| L.Log Transport Spending | | | | | | -0.434 *** |
| | | | | | | (0.0519) |
| Observations | 48474 | 49396 | 1852 | 47412 | 53189 | 43051 |
| Lags | 3-7 | 5-9 | 4-8 | 6-10 | 3-7 | 6-10 |

Arellano-Bond estimation of log expenditures in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The expenditures are estimated as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996. IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996, and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996, and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 17: **Arellano-Bond - Schooling Inputs (All Schools)**

| | (1) Ln teachers | (2) Teachers' ed | (3) Ln rooms (avail) | (4) Ln rooms (used) | (5) Ex. cap. in perc |
|---|---|---|---|---|---|
| Log FPM transfer | 0.0308*** | 0.0266** | 0.0164** | -0.0129* | 0.0657*** |
| | (0.00809) | (0.0129) | (0.00785) | (0.00684) | (0.00523) |
| L.Ln teachers | 0.866*** | | | | |
| | (0.0274) | | | | |
| L.Teachers' ed | | 0.805*** | | | |
| | | (0.00753) | | | |
| L.Ln rooms (avail) | | | 0.663*** | | |
| | | | (0.0506) | | |
| L.Ln rooms (used) | | | | 0.618*** | |
| | | | | (0.0463) | |
| L.Ex. cap. in perc | | | | | 0.559*** |
| | | | | | (0.0391) |
| N | 43048 | 37681 | 47873 | 47874 | 47873 |
| Lags | 3-7 | 3-7 | 4-8 | 3-7 | 5-9 |

Arellano-Bond estimation of five educational inputs in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. IV-style instruments are the lagged log of the state population coefficient and year dummies, while lagged values of the FPM transfer and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected clusterÐrobust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Table 18: **Arellano-Bond with GDP/Capita Interaction - Schooling Inputs (All Schools)**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Ln teachers | Teachers' ed | Ln rooms (avail) | Ln rooms (used) | Ex. cap. in perc |
| Log FPM transfer | 0.0631*** | -0.00649 | 0.00363 | -0.00812 | 0.0214*** |
| | (0.00830) | (0.0128) | (0.00682) | (0.00655) | (0.00370) |
| FPM * GDP | -0.00949* | 0.00552 | 0.00217 | -0.00397 | 0.0106*** |
| | (0.00520) | (0.00555) | (0.00412) | (0.00418) | (0.00208) |
| L.Ln teachers | 0.567*** | | | | |
| | (0.0414) | | | | |
| L.Teachers' ed | | 0.834*** | | | |
| | | (0.00922) | | | |
| L.Ln rooms (avail) | | | 0.612*** | | |
| | | | (0.0467) | | |
| L.Ln rooms (used) | | | | 0.625*** | |
| | | | | (0.0382) | |
| L.Ex. cap. in perc | | | | | 0.406*** |
| | | | | | (0.0570) |
| N | 38569 | 33736 | 43394 | 43395 | 43394 |
| Lags | 3-7 | 3-7 | 4-8 | 3-7 | 5-9 |

Arellano-Bond estimation of five educational inputs in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996. IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996, and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996, and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 19: **Arellano-Bond - Number of Teachers (in Natural Logs) - Schools by Level**

| | All | | Secondary | | Primary 5-8 | | Primary 1-4 | | Preschool | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Log FPM transfer | 0.0308*** | 0.0631*** | 0.0578* | 0.148*** | 0.106*** | 0.130*** | 0.0923** | -0.157*** | -0.0397 | -0.0121 |
| | (0.00809) | (0.00830) | (0.0312) | (0.0253) | (0.0297) | (0.0309) | (0.0427) | (0.0376) | (0.0392) | (0.0321) |
| FPM * GDP | | -0.00949* | | -0.0399*** | | -0.0724*** | | 0.245*** | | 0.00717 |
| | | (0.00520) | | (0.0144) | | (0.0204) | | (0.0322) | | (0.0176) |
| L.Ln teachers | 0.866*** | 0.567*** | | | | | | | | |
| | (0.0274) | (0.0414) | | | | | | | | |
| L.Ln teachers | | | 0.504*** | 0.635*** | | | | | | |
| | | | (0.0483) | (0.0287) | | | | | | |
| L.Ln teachers | | | | | 0.794*** | 0.721*** | | | | |
| | | | | | (0.0244) | (0.0287) | | | | |
| L.Ln teachers | | | | | | | 0.627*** | 0.570*** | | |
| | | | | | | | (0.0551) | (0.0553) | | |
| L.Ln teachers | | | | | | | | | 0.659*** | 0.650*** |
| | | | | | | | | | (0.0338) | (0.0347) |
| N | 43048 | 38569 | 39593 | 36417 | 34209 | 30804 | 35587 | 32104 | 40251 | 36463 |
| Lags | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 |

Arellano-Bond estimation of logged number of teachers at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 20: **Arellano-Bond - Number of Teachers (in Natural Logs) - Schools by Administration**

| | All | | State | | Municipal | | Private | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log FPM transfer | 0.0308*** | 0.0631*** | 0.0276* | 0.0660*** | 0.0895*** | 0.0712*** | -0.140** | -0.0750** |
| | (0.00809) | (0.00830) | (0.0150) | (0.0152) | (0.0128) | (0.0118) | (0.0567) | (0.0367) |
| FPM * GDP | | -0.00949* | | -0.0538*** | | 0.0224*** | | 0.00336 |
| | | (0.00520) | | (0.00979) | | (0.00665) | | (0.0197) |
| L.Ln teachers | 0.866*** | 0.567*** | | | | | | |
| | (0.0274) | (0.0414) | | | | | | |
| L.Ln teachers | | | 0.708*** | 0.559*** | | | | |
| | | | (0.0294) | (0.0314) | | | | |
| L.Ln teachers | | | | | 0.677*** | 0.630*** | | |
| | | | | | (0.0344) | (0.0313) | | |
| L.Ln teachers | | | | | | | 0.598*** | 0.632*** |
| | | | | | | | (0.0351) | (0.0310) |
| N | 43048 | 38569 | 42030 | 38126 | 42877 | 38434 | 22152 | 21574 |
| Lags | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 |

Arellano-Bond estimation of logged number of teachers at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 21: **Arellano-Bond - Average Teacher Qualification - Schools by Level**

| | All | | Secondary | | Primary 5-8 | | Primary 1-4 | | Preschool | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Log FPM transfer | 0.0266** | -0.00649 | 0.0335 | 0.00572 | 0.0456*** | 0.00568 | -0.00938 | -0.0303 | 0.0150 | 0.00304 |
| | (0.0129) | (0.0128) | (0.0208) | (0.0180) | (0.0169) | (0.0150) | (0.0303) | (0.0252) | (0.0300) | (0.0282) |
| FPM * GDP | | 0.00552 | | 0.00283 | | 0.00489 | | 0.00896 | | -0.000827 |
| | | (0.00555) | | (0.00839) | | (0.00969) | | (0.0148) | | (0.0159) |
| L.Teachers' ed | 0.805*** | 0.834*** | | | | | | | | |
| | (0.00753) | (0.00922) | | | | | | | | |
| L.Teachers' ed | | | 0.869*** | 0.855*** | | | | | | |
| | | | (0.0294) | (0.0281) | | | | | | |
| L.Teachers' ed | | | | | 0.895*** | 0.931*** | | | | |
| | | | | | (0.0336) | (0.0392) | | | | |
| L.Teachers' ed | | | | | | | 0.675*** | 0.647*** | | |
| | | | | | | | (0.0214) | (0.0225) | | |
| L.Teachers' ed | | | | | | | | | 0.644*** | 0.705*** |
| | | | | | | | | | (0.0230) | (0.0271) |
| N | 37681 | 33736 | 35154 | 32180 | 30224 | 27175 | 30899 | 27860 | 35578 | 32174 |
| Lags | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 |

Arellano-Bond estimation of average teacher qualification at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 22: **Arellano-Bond - Average Teacher Qualification - Schools by Administration**

| | All | | State | | Municipal | | Private | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log FPM transfer | 0.0266** | -0.00649 | 0.0353** | 0.0209 | 0.0148 | -0.0190 | -0.0545 | 0.00462 |
| | (0.0129) | (0.0128) | (0.0148) | (0.0139) | (0.0205) | (0.0178) | (0.0386) | (0.0317) |
| FPM * GDP | | 0.00552 | | 0.000529 | | 0.0206** | | -0.0280 |
| | | (0.00555) | | (0.00701) | | (0.00912) | | (0.0184) |
| L.Teachers' ed | 0.805*** | 0.834*** | | | | | | |
| | (0.00753) | (0.00922) | | | | | | |
| L.Teachers' ed | | | 0.942*** | 0.931*** | | | | |
| | | | (0.0212) | (0.0173) | | | | |
| L.Teachers' ed | | | | | 0.722*** | 0.752*** | | |
| | | | | | (0.0105) | (0.0129) | | |
| L.Teachers' ed | | | | | | | 0.534*** | 0.626*** |
| | | | | | | | (0.0975) | (0.0827) |
| N | 37681 | 33736 | 36849 | 33375 | 37572 | 33652 | 19806 | 19276 |
| Lags | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 |

Arellano-Bond estimation of average teacher qualification at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 23: **Arellano-Bond - Existing Classrooms (in Natural Logs) - Schools by Level**

| | All | | Secondary | | Primary 5-8 | | Primary 1-4 | | Preschool | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Log FPM transfer | 0.0164** | 0.00363 | -0.0888*** | 0.0745*** | -0.118*** | -0.0379 | 0.464*** | -0.0295 | -0.312*** | -0.205*** |
| | (0.00785) | (0.00682) | (0.0209) | (0.0224) | (0.0267) | (0.0285) | (0.0531) | (0.0447) | (0.0356) | (0.0348) |
| FPM * GDP | | 0.00217 | | -0.0521*** | | -0.0852*** | | 0.384*** | | -0.0445** |
| | | (0.00412) | | (0.0130) | | (0.0190) | | (0.0322) | | (0.0195) |
| L.Ln rooms (avail) | 0.663*** | 0.612*** | | | | | | | | |
| | (0.0506) | (0.0467) | | | | | | | | |
| L.Ln rooms (avail) | | | 0.170*** | 0.494*** | | | | | | |
| | | | (0.0607) | (0.0427) | | | | | | |
| L.Ln rooms (avail) | | | | | 0.819*** | 0.805*** | | | | |
| | | | | | (0.0273) | (0.0277) | | | | |
| L.Ln rooms (avail) | | | | | | | 0.0854 | 0.0933* | | |
| | | | | | | | (0.0658) | (0.0548) | | |
| L.Ln rooms (avail) | | | | | | | | | 0.459*** | 0.444*** |
| | | | | | | | | | (0.0394) | (0.0398) |
| N | 47873 | 43394 | 43907 | 40731 | 37760 | 34355 | 39804 | 36321 | 44510 | 40722 |
| Lags | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 |

Arellano-Bond estimation of logged existing classrooms at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 24: **Arellano-Bond - Existing Classrooms (in Natural Logs) - Schools by Administration**

| | All | | State | | Municipal | | Private | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log FPM transfer | 0.0164** | 0.00363 | 0.0540*** | 0.0589*** | -0.0186 | -0.0270** | 0.0754** | -0.00221 |
| | (0.00785) | (0.00682) | (0.0169) | (0.0144) | (0.0132) | (0.0118) | (0.0344) | (0.0313) |
| FPM * GDP | | 0.00217 | | 0.00598 | | 0.00665 | | 0.0352* |
| | | (0.00412) | | (0.00800) | | (0.00646) | | (0.0185) |
| L.Ln rooms (avail) | 0.663*** | 0.612*** | | | | | | |
| | (0.0506) | (0.0467) | | | | | | |
| L.Ln rooms (avail) | | | 0.748*** | 0.668*** | | | | |
| | | | (0.0346) | (0.0362) | | | | |
| L.Ln rooms (avail) | | | | | 0.652*** | 0.578*** | | |
| | | | | | (0.0360) | (0.0284) | | |
| L.Ln rooms (avail) | | | | | | | 0.525*** | 0.602*** |
| | | | | | | | (0.0503) | (0.0369) |
| N | 47873 | 43394 | 46775 | 42871 | 47580 | 43137 | 24432 | 23851 |
| Lags | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 | 4-8 |

Arellano-Bond estimation of logged number of existing classrooms at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 25: **Arellano-Bond - Used Classrooms (in Natural Logs) - Schools by Level**

| | All | | Secondary | | Primary 5-8 | | Primary 1-4 | | Preschool | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Log FPM transfer | -0.0129* | -0.00812 | -0.0348 | 0.0845*** | -0.128*** | 0.0116 | 0.311*** | -0.0629 | -0.370*** | -0.225*** |
| | (0.00684) | (0.00655) | (0.0222) | (0.0206) | (0.0252) | (0.0274) | (0.0371) | (0.0421) | (0.0352) | (0.0343) |
| FPM * GDP | | -0.00397 | | -0.0541*** | | -0.120*** | | 0.313*** | | -0.0761*** |
| | | (0.00418) | | (0.0124) | | (0.0214) | | (0.0313) | | (0.0175) |
| L.Ln rooms (used) | 0.618*** | 0.625*** | | | | | | | | |
| | (0.0463) | (0.0382) | | | | | | | | |
| L.Ln rooms (used) | | | 0.538*** | 0.626*** | | | | | | |
| | | | (0.0498) | (0.0332) | | | | | | |
| L.Ln rooms (used) | | | | | 0.808*** | 0.736*** | | | | |
| | | | | | (0.0238) | (0.0295) | | | | |
| L.Ln rooms (used) | | | | | | | 0.215*** | 0.134*** | | |
| | | | | | | | (0.0549) | (0.0514) | | |
| L.Ln rooms (used) | | | | | | | | | 0.530*** | 0.550*** |
| | | | | | | | | | (0.0296) | (0.0275) |
| N | 47874 | 43395 | 43907 | 40731 | 37760 | 34355 | 39805 | 36322 | 44510 | 40722 |
| Lags | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 |

Arellano-Bond estimation of logged existing classrooms at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 26: **Arellano-Bond - Used Classrooms (in Natural Logs) - Schools by Administration**

| | All | | State | | Municipal | | Private | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log FPM transfer | -0.0129* | -0.00812 | 0.0212 | 0.0581*** | -0.0527*** | -0.0351*** | 0.0320 | -0.0252 |
| | (0.00684) | (0.00655) | (0.0143) | (0.0142) | (0.0129) | (0.0113) | (0.0350) | (0.0302) |
| FPM * GDP | | -0.00397 | | -0.0128 | | 0.000610 | | 0.0175 |
| | | (0.00418) | | (0.00799) | | (0.00643) | | (0.0180) |
| [1em] L.Ln rooms (used) | 0.618*** | 0.625*** | | | | | | |
| | (0.0463) | (0.0382) | | | | | | |
| L.Ln rooms (used) | | | 0.771*** | 0.610*** | | | | |
| | | | (0.0265) | (0.0318) | | | | |
| L.Ln rooms (used) | | | | | 0.626*** | 0.620*** | | |
| | | | | | (0.0288) | (0.0232) | | |
| L.Ln rooms (used) | | | | | | | 0.590*** | 0.627*** |
| | | | | | | | (0.0384) | (0.0299) |
| N | 47874 | 43395 | 46775 | 42871 | 47581 | 43138 | 24432 | 23851 |
| Lags | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 | 3-7 |

Arellano-Bond estimation of logged number of used classrooms at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 27: **Arellano-Bond - Excess Classroom Capacity (in Percent of Existing Classrooms) - Schools by Level**

| | All | | Secondary | | Primary 5-8 | | Primary 1-4 | | Preschool | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Log FPM transfer | 0.0657*** | 0.0214*** | 0.0459*** | 0.00277 | 0.0456*** | 0.00381 | 0.0604*** | -0.0143** | 0.0422*** | 0.0177*** |
| | (0.00523) | (0.00370) | (0.0111) | (0.00763) | (0.00842) | (0.00563) | (0.00878) | (0.00707) | (0.00657) | (0.00501) |
| FPM * GDP | | 0.0106*** | | 0.0145*** | | 0.0135*** | | 0.0196*** | | 0.00228 |
| | | (0.00208) | | (0.00395) | | (0.00351) | | (0.00441) | | (0.00313) |
| L.Ex. cap. in perc | 0.559*** | 0.406*** | | | | | | | | |
| | (0.0391) | (0.0570) | | | | | | | | |
| L.Ex. cap. in perc | | | 0.482*** | 0.327*** | | | | | | |
| | | | (0.0578) | (0.0630) | | | | | | |
| L.Ex. cap. in perc | | | | | 0.580*** | 0.314*** | | | | |
| | | | | | (0.0611) | (0.0801) | | | | |
| L.Ex. cap. in perc | | | | | | | 0.234** | -0.251*** | | |
| | | | | | | | (0.103) | (0.0782) | | |
| L.Ex. cap. in perc | | | | | | | | | 0.412*** | 0.328*** |
| | | | | | | | | | (0.0817) | (0.0767) |
| N | 47873 | 43394 | 43907 | 40731 | 37760 | 34355 | 39804 | 36321 | 44510 | 40722 |
| Lags | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 |

Arellano-Bond estimation of excess classroom capacity at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

Table 28: **Arellano-Bond - Excess Classroom Capacity (in Percent of Existing Classrooms) - Schools by Administration**

| | All | | State | | Municipal | | Private | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log FPM transfer | 0.0657*** | 0.0214*** | 0.0641*** | 0.0180*** | 0.0621*** | 0.0165*** | 0.0346** | -0.00705 |
| | (0.00523) | (0.00370) | (0.00788) | (0.00581) | (0.00575) | (0.00437) | (0.0154) | (0.0105) |
| FPM * GDP | | 0.0106*** | | 0.0148*** | | 0.00844*** | | 0.0215*** |
| | | (0.00208) | | (0.00322) | | (0.00251) | | (0.00619) |
| [1em] L.Ex. cap. in perc | 0.559*** | 0.406*** | | | | | | |
| | (0.0391) | (0.0570) | | | | | | |
| L.Ex. cap. in perc | | | 0.554*** | 0.383*** | | | | |
| | | | (0.0606) | (0.0689) | | | | |
| L.Ex. cap. in perc | | | | | 0.464*** | 0.278*** | | |
| | | | | | (0.0577) | (0.0748) | | |
| L.Ex. cap. in perc | | | | | | | 0.501*** | 0.242** |
| | | | | | | | (0.0890) | (0.102) |
| N | 47873 | 43394 | 46775 | 42871 | 47580 | 43137 | 24432 | 23851 |
| Lags | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 | 5-9 |

Arellano-Bond estimation of excess classroom capacity at type-j schools in municipality i and state s at time t, estimated by two-step difference GMM with deviations in first differences, using xtabond2. The dependent variables are estimated as a function of logged transfers (odd-numbered columns) and as a function of logged transfers and logged transfers interacted with log municipal GDP/capita in 1996 (even-numbered columns). IV-style instruments are the lagged log of the state population coefficient, this variable interacted with log municipal GDP/capita in 1996 (in even columns), and year dummies. Lagged values of the FPM transfer, the interaction of FPM transfer and log municipal GDP/capita in 1996 (in even columns), and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.

# A Data Appendix

## A.1 Derivation of Variables

The following subsections describe the derivation of our outcome variables. We consider three types of school resources (number of classrooms, number of teachers, and the education level of teachers), and a series of dummy variables capturing educational offerings within the municipality.

### A.1.1 Classrooms

There are four categories of classrooms reported in the *Censo escolar*: permanent, temporary, internal and external (see Table 29). In the survey document itself, these four variables come from two separate questions. In the first, schools are asked to report the number of existing classrooms, with separate blanks provided for permanent and temporary rooms. In the second, school are asked the number of used classrooms, with separate blanks for within the establishment and outside the establishment.[29] The mean number of used classrooms (internal plus external) is slightly below the mean number of existing classrooms (permanent plus temporary).

Both of these measures are of interest. The first is the most appropriate measure of physical plant: the number of classrooms available in the municipality. The second, however, gives us a measure of intensity of use. An increase in the number of classrooms used is perhaps a more convincing signal of education provision than an increase in the number of classrooms per se – particularly when most municipalities display some degree of excess capacity.

To complement our study of these two variables we create a third measure which captures the difference between the two. This variable, the excess capacity in percent, is calculated as (existing classrooms-used classrooms)/existing classrooms. These three measures are summarised in Table 30.

[29]This example is taken from the 1999 census form (translation by the authors). The original wording of the questions is: *Número de Salas de Aula Existentes,* with options *Permanentes* and *Provisórias*; and *Número de Salas de Aula Utilizadas,* with options *No Estabelecimento* and *Fora do Estabelecimento*.

#### A.1.2 Teachers

The total number of active teachers is reported directly by schools in the *Censo escolar*.[30] While the number of teachers is also reported according to the level at which they teach, it is possible that some teachers would teach at more than one level. We therefore use the direct count reported in the survey. Table 31 lists the average number of active teachers by municipality for the duration of our panel.

#### A.1.3 Teachers' Education

To measure teacher qualification level, we collapse a larger list of education levels into an index from 0-3. To build this index, we define four levels of education, described in Table 32. We code each teacher according this index, and average the scores over the municipality.[31] The average number of teachers at each level is summarised in Table 33. Table 34 demonstrates the steady improvements in teacher numbers and qualifications over the course of the sample (note that details on teacher education were not collected in 1997).

### A.2 Coding Schools by Level of Instruction

In the sub-analysis in Section ?? we consider the effect of changes in transfers on schools offering classes at different levels of instruction. Our interest is primarily to see whether investments are being made particularly towards the provision of levels of education which may not yet be widely available: specifically upper levels, but also to some extent preschool. Many schools offer multiple levels of instruction, making it difficult to cleanly attribute growth to one level of instruction or another. Since we are most interested in potential expansions towards the 'extremities' of the grade distribution, we attribute schools to the most restrictive grades they offer.

Specifically, we first code all schools that offer secondary school classes as secondary schools. Next, any school that offers senior primary grades 5-8, but does not offer secondary school classes, is coded as a 5-8 school. Because junior primary schools, offering grades 1-4 are the most

[30]From the 1999 census questionnaire: *Total de Professores em Exercício (em sala de aula).*

[31]In the data, teacher education is listed according the level at which they teach; in other words, for each level of instruction provided at the school, the number of teachers with different levels of completed education is reported. This introduces some error into our measurement, since some teachers may be teaching at different levels of instruction.

common, we next define preschools. We code preschools as any school that offers preschool, but not secondary or junior primary classes. Finally, the remaining schools, coded as junior primary schools, are those schools which offer grades 1-4 and *do not* offer classes at any of the other levels under consideration.

This mapping is not perfect. Secondary schools often co-exist with senior primary schools: of the 20,229 secondary schools in the 2001 census, 16,869 also offered senior primary grades. Even more commonly, senior primary schools also offer junior primary grades: of the 34,222 senior primary schools in the 2001 census, 30,030 offered junior primary classes, and 15,487 even offered preschool. On the younger end, 55,697 of the 87,845 preschools also offer junior primary grades. (This leaves 70,736 schools offering exclusively junior primary classes.)

It is important to keep in mind that this mapping doesn't allow us to pin-point exactly to what level of education these investments were targeted. An expansion in the number of classrooms in schools offering secondary grades does not necessarily mean an expansion in the number of classes for secondary students, if those schools also offer junior primary grades. Nevertheless, given our interest in measuring improvements in local education supply, this coding is useful: if schools offering secondary classes have more resources, it is likely that they will be able to use them to reduce bottlenecks in the education system.

# B Extensions

## B.1 Bounding Arellano-Bond Autoregressive Coefficients

As described in Bond (2002 [8]), a comparison of two mis-specification of the empirical model can provide a useful check on the consistency of Arellano-Bond estimates. If the true model is a simple AR(1) process with a time-invariant individual effect, then standard OLS estimates will tend to be biased upwards, while estimates from a fixed-effects estimation will tend to be biased downwards.

While these parameters are not the focus of our study, we can nevertheless check whether the autoregressive coefficients we obtain are close to satisfying this condition. Tables 36 and 37 present the results of standard OLS and fixed-effects regressions which correspond to the second-stage finance estimates in Table 12. A comparison of autoregressive parameter estimates in the three tables shows that, by and large, our primary estimates fall nicely between the equivalent estimates in these two mis-specified cases. Notable exceptions are lagged current revenue, which in the fixed-effects regression has a coefficient well in excess of one (very imprecisely estimated), and lagged capital expenditure, which in our primary estimation has a negative coefficient.

Tables 38 and 39 present equivalent results for educational inputs, replicating those found in Table 17. Table 38 presents result from an OLS estimation (the upper bound), while Table 39 presents fixed-effects results (the lower bound). Our Arellano-Bond point estimates for the autoregressive parameter fall between these two for all five variables.

## Appendix: Figures and Tables

Table 29: **Summary Statistics: Number of Classrooms**

| Variable | Mean | Std. Dev. | N |
|---|---|---|---|
| Permanent | 122.729 | 962.546 | 58738 |
| Temporary | 9.808 | 17.522 | 58738 |
| Internal | 113.708 | 127.167 | 58738 |
| External | 7.751 | 13.952 | 58738 |

*Source: Censo Escolar* 1996-2006.

Table 30: **Average Number of Classrooms per Municipality**

| Year | Existing | Used | Excess |
|---|---|---|---|
| 1996 | 121.322 | 114.7723 | 6.575325 |
| 1997 | 115.3933 | 108.2889 | 7.104322 |
| 1998 | 117.6337 | 112.4051 | 5.228619 |
| 1999 | 122.2651 | 114.4248 | 7.84032 |
| 2000 | 125.9148 | 118.0751 | 7.839672 |
| 2001 | 129.3532 | 121.9147 | 7.438493 |
| 2002 | 131.9471 | 124.9769 | 6.97024 |
| 2003 | 134.2318 | 127.0407 | 7.191084 |
| 2004 | 136.2203 | 128.7431 | 7.477248 |
| 2005 | 138.3223 | 130.7539 | 7.568413 |
| 2006 | 140.9394 | 133.628 | 7.311481 |

*Source: Censo Escolar* 1996-2006.

Table 31: **Average Number of Teachers per Municipality**

| Year | Mean | Std. Dev. | N |
|---|---|---|---|
| 1997 | 214.6539 | 262.2019 | 5368 |
| 1998 | 218.2594 | 266.8445 | 5367 |
| 1999 | 228.6816 | 277.8041 | 5367 |
| 2000 | 240.8315 | 290.1278 | 5364 |
| 2001 | 248.6935 | 300.4876 | 5413 |
| 2002 | 256.6168 | 307.55 | 5410 |
| 2003 | 265.0129 | 314.406 | 5406 |
| 2004 | 269.1092 | 320.0275 | 5402 |
| 2005 | 271.9815 | 320.4854 | 5401 |
| 2006 | 278.5357 | 326.6278 | 5400 |

*Source: Censo escolar* 1996-2006.

Table 32: **Teacher's Education Index**

| Score | Corresponding education levels |
|---|---|
| 0 | Less than primary completion |
| 1 | Completed primary school |
| 2 | Completed secondary school |
| 3 | Any tertiary degree |

Definitions by the authors.

Table 33: **Summary Statistics: Teacher Education**

| Variable | Mean | Std. Dev. | N |
|---|---|---|---|
| Level 0 | 6.115 | 20.118 | 53370 |
| Level 1 | 8.85 | 19.594 | 53370 |
| Level 2 | 134.552 | 160.154 | 53370 |
| Level 3 | 139.908 | 388.847 | 53370 |

*Source: Censo escolar* 1996-2006.

Table 34: **Evolution of Teacher Qualifications**

| Year | Level 0 | Level 1 | Level 2 | Level 3 |
|---|---|---|---|---|
| 1996 | 20.23657 | 16.64773 | 126.0438 | 110.5802 |
| 1997 | . | . | . | . |
| 1998 | 11.88131 | 12.7438 | 118.1468 | 96.57947 |
| 1999 | 9.113099 | 12.23905 | 129.4196 | 103.7654 |
| 2000 | 6.726324 | 12.29008 | 137.0677 | 111.9314 |
| 2001 | 3.984851 | 10.43063 | 144.9714 | 121.0031 |
| 2002 | 2.031054 | 5.808318 | 150.3573 | 133.4375 |
| 2003 | 1.400111 | 3.789678 | 147.1966 | 150.0906 |
| 2004 | 2.686783 | 5.32562 | 143.5837 | 167.5085 |
| 2005 | 2.423996 | 5.285503 | 130.132 | 188.1972 |
| 2006 | 2.202963 | 4.82963 | 117.5313 | 212.3009 |

*Source: Censo escolar* 1996-2006.

Table 35: **Municipal Mean Number of Schools by Level of Instruction Offered**

| Year | Preschool | Jr. Primary | Sr. Primary | Secondary |
|---|---|---|---|---|
| 1996 | 17.980 | 33.531 | 5.555 | 1.886 |
| 1997 | 11.235 | 30.115 | 5.003 | 1.849 |
| 1998 | 13.327 | 28.219 | 5.277 | 1.959 |
| 1999 | 13.969 | 27.321 | 5.641 | 1.774 |
| 2000 | 14.445 | 26.831 | 6.033 | 2.121 |
| 2001 | 14.943 | 25.701 | 6.316 | 2.183 |
| 2002 | 14.797 | 24.618 | 6.658 | 2.288 |
| 2003 | 14.872 | 23.836 | 6.928 | 2.349 |
| 2004 | 14.530 | 23.289 | 7.126 | 2.478 |
| 2005 | 15.102 | 22.519 | 7.242 | 2.522 |
| 2006 | 15.380 | 21.771 | 7.389 | 2.608 |

*Source: Censo escolar*, 1996-2006.

Table 36: **Finance: OLS with Lagged Dependent Variable (Should be Upward Biased)**

| | (1) Log Current Revenue | (2) Log Property Tax Revenue | (3) Log ISS Tax Revenue | (4) Log Current Expenditure | (5) Log Capital Expenditure | (6) Log Education Spending |
|---|---|---|---|---|---|---|
| Log FPM transfer | -0.238*** | -1.294*** | -0.181*** | -0.314*** | 0.436*** | -0.262*** |
| | (0.0321) | (0.242) | (0.0628) | (0.0489) | (0.107) | (0.0280) |
| L.Log Current Revenue - Municipal | 1.113*** | | | | | |
| | (0.0176) | | | | | |
| L.Log Tax revenue - property tax - Municipal | | 1.088*** | | | | |
| | | (0.0296) | | | | |
| L.Log Tax revenue - ISS - Municipal | | | 0.912*** | | | |
| | | | (0.0113) | | | |
| L.Log Current Expenditure | | | | 1.141*** | | |
| | | | | (0.0265) | | |
| L.Log Capital expenditure - Municipal | | | | | 0.620*** | |
| | | | | | (0.0339) | |
| L.Log Expenses by function - education and culture - Municipal | | | | | | 1.085*** |
| | | | | | | (0.0154) |
| Constant | 1.924*** | 19.49*** | 4.086*** | 2.706*** | -1.301 | 2.866*** |
| | (0.218) | (3.495) | (0.855) | (0.344) | (1.197) | (0.231) |
| Observations | 75470 | 70419 | 73949 | 79835 | 69908 | 66100 |

OLS estimation of log revenue and spending in municipality i and state s at time t, estimated as a function of the log transfer to the municipality. The transfer is instrument by lagged log of the state population coefficient. Time dummies are included. The results are expected to be upwards biased. Standard errors clustered at the municipal level are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Table 37: **Finance: Fixed-Effect with Lagged Dependent Variable (Should be Downward Biased)**

| | (1)<br>Log Current Revenue | (2)<br>Log Property Tax Revenue | (3)<br>Log ISS Tax Revenue | (4)<br>Log Current Expenditure | (5)<br>Log Capital Expenditure | (6)<br>Log Education Spending |
|---|---|---|---|---|---|---|
| Log FPM transfer | -17.19 | -3.093 | 71.97 | 0.533*** | 0.840*** | 1.209*** |
| | (15.39) | (4.634) | (120.9) | (0.0105) | (0.0248) | (0.0159) |
| L.Log Current Revenue - Municipal | 2.701 | | | | | |
| | (1.874) | | | | | |
| L.Log Tax revenue - property tax - Municipal | | 0.284*** | | | | |
| | | (0.00428) | | | | |
| L.Log Tax revenue - ISS - Municipal | | | 0.676** | | | |
| | | | (0.339) | | | |
| L.Log Current Expenditure | | | | 0.479*** | | |
| | | | | (0.00448) | | |
| L.Log Capital expenditure - Municipal | | | | | 0.240*** | |
| | | | | | (0.00392) | |
| L.Log Expenses by function - education and culture - Municipal | | | | | | 0.298*** |
| | | | | | | (0.00543) |
| Constant | 242.8 | 56.88 | -1131.8 | 0.119 | -2.133*** | -8.132*** |
| | (211.4) | (73.21) | (1912.4) | (0.116) | (0.377) | (0.188) |
| Observations | 75470 | 70419 | 73949 | 79835 | 69908 | 66100 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Log Current Revenue | Log Property Tax Revenue | Log ISS Tax Revenue | Log Current Expenditure | Log Capital Expenditure | Log Education Spending |

Fixed-Effects estimation of log revenue and spending in municipality i and state s at time t, estimated as a function of the log transfer to the municipality. The transfer is instrument by lagged log of the state population coefficient. Time dummies are included. The results are expected to be upwards biased. Standard errors are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Table 38: **School Inputs: OLS with Lagged Dependent Variable (Should be Upward Biased)**

| | (1) Ln teachers | (2) Teachers' ed | (3) Ln rooms (avail) | (4) Ln rooms (used) | (5) Ex. cap. in perc |
|---|---|---|---|---|---|
| Log FPM transfer | -0.0129*** | 0.0380** | -0.000298 | -0.0187*** | 0.0564*** |
| | (0.00388) | (0.0162) | (0.00410) | (0.00381) | (0.00599) |
| L.Ln teachers | 1.001*** | | | | |
| | (0.00173) | | | | |
| L.Teachers' ed | | 0.837*** | | | |
| | | (0.00505) | | | |
| L.Ln rooms (avail) | | | 0.995*** | | |
| | | | (0.00188) | | |
| L.Ln rooms (used) | | | | 1.003*** | |
| | | | | (0.00174) | |
| L.Ex. cap. in perc | | | | | 0.616*** |
| | | | | | (0.0109) |
| Constant | 0.210*** | -0.168 | 0.0544 | 0.280*** | -0.848*** |
| | (0.0499) | (0.233) | (0.0553) | (0.0508) | (0.0925) |
| Observations | 48472 | 43105 | 53298 | 53299 | 53298 |

OLS estimation of school inputs in municipality i and state s at time t, estimated as a function of the log transfer to the municipality. The transfer is instrument by lagged log of the state population coefficient. Time dummies are included. The results are expected to be upwards biased. Standard errors clustered at the municipal level are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Table 39: **School Inputs: Fixed-Effect with Lagged Dependent Variable (Should be Downward Biased)**

| | (1) Ln teachers | (2) Teachers' ed | (3) Ln rooms (avail) | (4) Ln rooms (used) | (5) Ex. cap. in perc |
|---|---|---|---|---|---|
| Log FPM transfer | $0.0690^{***}$ | $0.132^{***}$ | $0.389^{***}$ | $0.533^{***}$ | $-0.223^{***}$ |
| | (0.0111) | (0.0229) | (0.0448) | (0.0522) | (0.0277) |
| L.Ln teachers | $0.512^{***}$ | | | | |
| | (0.00435) | | | | |
| L.Teachers' ed | | $0.581^{***}$ | | | |
| | | (0.00527) | | | |
| L.Ln rooms (avail) | | | $0.465^{***}$ | | |
| | | | (0.00872) | | |
| L.Ln rooms (used) | | | | $0.394^{***}$ | |
| | | | | (0.0114) | |
| L.Ex. cap. in perc | | | | | $0.181^{***}$ |
| | | | | | (0.00714) |
| Constant | $1.453^{***}$ | $-0.937^{***}$ | $-3.652^{***}$ | $-5.633^{***}$ | $3.531^{***}$ |
| | (0.164) | (0.349) | (0.666) | (0.770) | (0.433) |
| Observations | 48472 | 43105 | 53298 | 53299 | 53298 |

Fixed-Effects estimation of school inputs in municipality i and state s at time t, estimated as a function of the log transfer to the municipality. The transfer is instrument by lagged log of the state population coefficient. Time dummies are included. The results are expected to be upwards biased. Standard errors are in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level respectively.

Table 40: **Arellano-Bond System GMM Estimation**

| | (1)<br>Log Current Revenue | (2)<br>Log Property Tax Revenue | (3)<br>Log ISS Tax Revenue | (4)<br>Log Current Expenditure | (5)<br>Log Capital Expenditure | (6)<br>Log Education Spending |
|---|---|---|---|---|---|---|
| Log FPM transfer | 0.269*** | -0.0786 | 0.193*** | 0.0864*** | 1.213*** | 0.0413** |
| | (0.0161) | (0.0483) | (0.0438) | (0.0103) | (0.0424) | (0.0201) |
| L.Log Current Revenue | 0.764*** | | | | | |
| | (0.0105) | | | | | |
| L.Log Property Tax Revenue | | 1.006*** | | | | |
| | | (0.00453) | | | | |
| L.Log ISS Tax Revenue | | | 0.843*** | | | |
| | | | (0.0151) | | | |
| L.Log Current Expenditure | | | | 0.883*** | | |
| | | | | (0.00944) | | |
| L.Log Capital Expenditure | | | | | -0.103*** | |
| | | | | | (0.0320) | |
| L.Log Education Spending | | | | | | 0.968*** |
| | | | | | | (0.0114) |
| Constant | -0.224 | 1.210 | -0.954 | 0.584*** | -3.030*** | -0.106 |
| | (0.168) | (0.745) | (0.627) | (0.0780) | (0.497) | (0.180) |
| Observations | 75470 | 70419 | 73949 | 79835 | 69908 | 66100 |
| Lags | 4-8 | 3-7 | 3-7 | 5-9 | 5-9 | 5-9 |

Arellano-Bond estimation of log revenues and expenditures in municipality i and state s at time t, estimated by two-step system GMM with deviations in first differences, using xtabond2. IV-style instruments are the lagged log of the state population coefficient and year dummies, while lagged values of the FPM transfer and the dependent variable are specified as GMM-style instruments (the included lags for each equation are specified in the table). Windmeijer-corrected cluster-robust errors are reported in parentheses. Stars indicate statistical significance at the 1% (***), 5% (**), and 10% (*) level, respectively.