# Heterogeneous Peer Effects in the Classroom

Michela M. Tincani[1]

September 10, 2018

[1]Tincani: University College London, 30 Gordon Street, London, WC1H0BE, UK, m.tincani@ucl.ac.uk. This is a substantially revised version of a previous working paper titled "Heterogeneous Peer Effects and Rank Concerns: Theory and Evidence", HCEO WP 2017-006 and CESifo WP 6331. This paper focuses on the empirical finding of heterogeneous peer effects. A separate working paper co-authored with Konrad Mierendorff focuses on rank concerns as a mechanism behind heterogeneous peer effects. I am particularly grateful to Orazio Attanasio, Ed Hopkins, Aureo de Paula, Petra Todd and Ken Wolpin for their feedback and encouragement. I would also like to thank Richard Blundell, Denis Chetverikov, Martin Cripps, Mariacristina De Nardi, Steven Durlauf, Christian Dustmann, Jan Eeckhout, Tatiana Kornienko, Dennis Kristensen, Stephen Machin, Costas Meghir, Chris Neilson, Lars Nesheim, Imran Rasul, Elie Tamer, Martin Weidner, Daniel Wilhelm and participants at various seminars and conferences for insightful comments and discussions. Julia Schmieder provided excellent research assistance. Research funding from the Centre for Microdata Methods and Practice and from the European Research Council's grant number IHKDC-249612 is gratefully acknowledged. I am grateful to the Chilean Ministry of Education and to the *Agencia de Calidad de la Educación* for access to some of the data used in this research. The views reported here are those of the author and do not necessarily reflect views of the Ministry or of the *Agencia*.

**Abstract**

This paper flexibly estimates peer effects in the classroom using a large administrative dataset of Chilean eighth graders, augmented with data from the propagation of the Chilean 2010 earthquake. It combines this natural experiment, that generated student-level shocks, with semi-parametric econometrics. Peer effects are identified from variation, across peer groups, in within-group shock realizations, keeping peer characteristics such as socio-economic status and lagged test scores constant. No restrictions on the shape of peer effects are imposed. The econometric model can be adapted to estimate the heterogeneous impacts of any moment of the distribution of peer characteristics. I use it to estimate the heterogeneous effects of changing the variance of peer characteristics, and show that an effort game in the classroom, in which achievement rank enters students' payoffs, can explain the estimated patterns.

# 1 Introduction

Peer effects have been widely studied by economists in many contexts, for example, career choices, health behaviors, crime and education. Typically, peer effect models describe an outcome of interest as a function of some feature of a peer group. The simplest model examines the importance of the mean of peer characteristics in shaping own behaviour. While this model seems to capture social influences in, for example, crime and risky health behaviors, it is not appropriate in other contexts.

In education, a growing body of empirical evidence points to the nonlinearity and heterogeneity of peer effects on test scores. Other moments beyond the mean of peer characteristics matter, and they do so differently for different students (see the surveys in Epple and Romano (2011) and Sacerdote (2014)). However, the empirical literature relies on parametric models that implicitly impose restrictions on the shape of peer effects. These restrictions may make it difficult to uncover the true shape of peer effects and, ultimately, the mechanisms that drive them.

This paper uses an econometric model for the estimation of peer effects that imposes fewer restrictions on the shape of peer effects than existing methods. It can detect any pattern of heterogeneity of peer effects across students. For identification, it does not rely on variation in peer characteristics such as the lagged test scores and socioeconomic status of peers. This is helpful, because it avoids the usual confounding influences that co-vary with these peer characteristics, and that have been the focus of much of the empirical literature on peer effects in education (for studies that use "natural exogenous" variation in peer characteristics, see, for example, Hoxby (2000), Angrist and Lang (2004), Hoxby and Weingarth (2005), Lavy, Paserman, and Schlosser (2012), Imberman, Kugler, and Sacerdote (2012); for studies that use controlled assignment to groups, see, for example, Sacerdote (2001), Zimmerman (2003), Duflo, Dupas, and Kremer (2011), Carrell, Sacerdote, and West (2013), Booij, Leuven, and Oosterbeek (2016) ).[1]

[1]There are at least three well understood challenges in the identification of peer effects. When the object of interest is, like in this paper, the effect of peer characteristics on own outcomes, also known as social contextual effects, there are two potential confounders. The first one are group level unobservables that co-vary with the characteristics of the group members. For example, in a school setting, different teachers may be assigned to classrooms with different kinds of students, and not all teacher characteristics are observed. The second one concerns self-selection of individuals into peer groups based on own unobservables and on the characteristics of the group members. These two challenges generate correlation between observable group members' characteristics and unobservable group or group members' characteristics that affect outcomes. In the terminology of Manski (1993), these two challenges are grouped into the term "correlated effects". Random or quasi-random assignment to peer groups helps in the identification of social contextual effects. The third identification challenge concerns the estimation of endogenous peer effects according to the terminology in Manski (1993), that is, the impact of peer

The econometric model has two key features. First, it characterises each student with a type that is a single index of a vector of characteristics which includes an "excluded element". That is, the self-selection of individuals into peer groups and unobserved peer group characteristics, such as teacher quality, do not depend on how this element is distributed in the peer group (or, at least, on the moment of this distribution that is of interest to the researcher). This means that varying the peer distribution (or the moment of interest) of this element does not give rise to correlated effects (Manski 1993). I use the local intensity of the Chilean 2010 earthquake at a student's home, which I can map into an expected level of structural damage to the home, as an excluded element.[2]

Second, the model uses semi-parametric techniques that allow me to be ex-ante agnostic about the shape of peer effects. In particular, it uses semi-parametric single-index models (Ichimura 1993) to flexibly estimate achievement as a function of student type in each peer group separately. It then preforms point-by-point comparisons of the estimated functions across peer groups that differ only in terms of the realizations of the excluded element of the vector of student characteristics. As a result, peer effects can be estimated as non-parametric functions of student characteristics.

Local earthquake intensity at a student's home plays the role of the excluded element of the vector of student characteristics. The self-selection of students into classrooms and the assignment of teachers and other resources to classrooms occurred before the earthquake struck and, therefore, cannot depend on the local intensity of the earthquake at students' homes. However, a different challenge to identification arises, because local earthquake intensity depends on the location of a student's home, which could correlate with unobserved student characteristics. Using a cohort of students that is observed before the earthquake struck and, therefore, was not affected by it, I find that there is a pre-existing correlation between geographic location of classmates and achievement. In the cohort that was affected by the earthquake, geographic location of classmates mechanically

actions on own actions. This is a simultaneous equations issue, akin to the identification of best response functions in games or to demand and supply functions when only market equilibria are observed. Random assignment to groups does not necessarily solve this issue, which can instead be addressed through exclusion restrictions coming through network structures (see, for example, Bramoullé, Djebbari, and Fortin (2009)). Blume, Brock, Durlauf, and Jayaraman (2015) provide a systematic analysis of identification of both types of peer effects in linear social interaction models.

[2]Natural disasters have been used before to identify peer effects in education, see, for example, Cipollone and Rosolia (2007), Imberman, Kugler, and Sacerdote (2012) and Sacerdote (2008). In contrast to previous studies, this paper does not use forced relocations of students for identification. This distinguishes this paper also from the experimental and quasi-experimental literatures that use variation in assignment to peer groups, e.g. dorms (Sacerdote 2001, Zimmerman 2003, Stinebrickner and Stinebrickner 2006, Kremer and Levy 2008, Garlick 2016) or classrooms (Duflo, Dupas, and Kremer 2011, Whitmore 2005, Kang 2007).

co-varies with their local earthquake intensities. Therefore, naive cross-sectional comparisons between post-earthquake classrooms that have different distributions of earthquake intensities would give rise to correlated effects (Manski 1993). To address this, I embed the semi-parametric method described above within a difference-in-differences framework. The resulting model is related to the nonlinear difference-in-differences models in, for example, Athey and Imbens (2006), Abadie (2005), Heckman, Ichimura, Smith, and Todd (1998), and Blundell, Costa Dias, Meghir, and Van Reenen (2004), but it addresses specific challenges of this context that those models are not well-suited to address.

Intuitively, I use a pre-earthquake cohort of students as a control group where correlated effects arise but effects due to damage to peers' homes do not. I measure correlated effects on this control group, and nett them out of the effects calculated on the treatment group, i.e., on the post-earthquake cohort of students. In practice, this adjustment means that I must perform double differences rather than single differences of functions: one within cohorts and one across cohorts. Intuitively, the identifying assumption of this differences-in-differences adjustment is that the correlation between geographic location of the student body and achievement is constant across cohorts in the absence of an earthquake. Using data not used in estimation on regions never affected by the earthquake, I test and do not reject this identifying assumption.

Like in many papers in the literature on peer effects in education, this paper has no ambition to separate endogenous and contextual/exogenous peer effects in the terminology of Manski (1993). The estimated peer effect functions are composite estimates that incorporate both. When interpreting the estimates, I allow for both types of peer effects. For example, I consider a reaction of teacher effort to changes in peer characteristics, which could be interpreted as a contextual peer effect, whereby peer characteristics (indirectly) affect own outcomes through the teacher's response. I also consider an effort game in the classroom. Comparative statics from the model compare the equilibrium distribution of achievement under different group configurations. The estimated patterns are interpreted as the outcome of student effort choices: changing the characteristics of peers has an effect on own effort choices through its effect on peers' effort choices.[3]

In terms of data construction, I combine a large administrative dataset on over 350,000 Chilean students with information on the Chilean 2010 earthquake. I use four waves of the SIMCE dataset (*Sistema de Medición de la Calidad de la Educación*, 2005, 2007, 2009,

[3]Notice that, even if this were the only source of peer effects in the data generating process, the empirical model would not identify the endogenous peer effect, that is, the best response function. Rather, it would trace how the equilibrium achievement function varies under different configurations of peer characteristics.

2011), with information on students, teachers, classrooms and schools. The outcomes of interest are standardised test scores in Mathematics and Spanish in the $8^{th}$ grade. School and classroom identifiers allow me to match students to classmates, teachers and schools, making this dataset useful to study peer effects at classroom level. I merge this educational data with a measure of local earthquake intensity at students' homes, which I obtained from the structural engineering literature. The measure is based on seismic intensity according to the Medvedev-Sponheuer-Karnik scale, and it can be mapped into a measure of damage to student's homes caused by the earthquake. The resulting dataset has three key features: it is longitudinal (test scores are observed twice for each student), it contains two cohorts (one, the post-earthquake cohort, affected by the earthquake in the second time period; and one, the pre-earthquake cohort, never affected), and it contains geographic variation in the intensity of exposure to the earthquake of each student.

Preliminary data analysis indicates that test score growth between the $4^{th}$ and the $8^{th}$ grade was 0.02 to 0.04 standard deviations lower for students affected by the earthquake compared to those who were not.[4] Additionally, students' test scores were affected differently depending on intensity of exposure. Moreover, the intensity of exposure to the earthquake of *peers* matters for *own* achievement. Exploiting variation in test score growth across cohorts and across geographic locations of classmates, I find that, keeping peer characteristics and other school and teacher characteristics constant, the average level of damages among a student's peers had insignificant or negligible effects, while the dispersion in peer damages had a significant and sizeable negative effect on own test scores.[5] Therefore, dispersion in classmates' damages is an empirically relevant margin in the context of my data. I exploit this margin to estimate the heterogeneous impacts of dispersion in peer characteristics.

Empirical findings indicate that student type is decreasing in lagged test score, parental education and parental income, its relationship with gender depends on the subject, and it is increasing in earthquake intensity at the point estimates, more so for females and low income students. Moreover, a post-estimation test cannot reject that achievement is monotonically decreasing in type. This first set of findings points to the interpretation of student type as a factor that decreases student productivity, like, for example, the cost of study effort. However, the analysis of treatment effects does not rely on a specific interpretation of student type.

[4] This is an estimate from a difference-in-difference regression model that compares trends in test score growth across cohorts between regions affected and not affected by the earthquake.

[5] This result holds irrespective of how dispersion is measured, e.g., standard deviation, coefficient of variation and various interquartile ranges.

The variance of peer types has heterogeneous impacts on student achievement. In particular, while the descriptive results based on linear models suggested a negative impact, the semi-parametric approach reveals that not everyone is hurt by dispersion in peer types. Some students at the tails of the type distribution benefit from an increase in the variance of peer types.[6] Moreover, who benefits varies across subjects.

I consider a number of mechanisms. First, I test whether the empirical findings can be explained by unobserved school level inputs. To do so, I derive a sufficient condition for the findings to not be driven by unobserved school inputs, and present evidence suggesting that this condition is satisfied. Intuitively, I show that school fixed effects did not react to variance in earthquake damages. Second, I measure teachers' effort in class through the fraction of the curriculum that the teacher was able to cover during the year, and, similarly, find no evidence that it reacted to the variance in earthquake damages in the classroom. While this analysis does not definitely rule out teacher effort as one of the drivers of impact, it suggests that other channels should be explored as well.

The third channel I explore is student effort. Following the approach adopted in Blume, Brock, Durlauf, and Jayaraman (2015), I propose a theoretical framework that micro-founds peer effects through an effort game in the classroom. Specifically, I show that a model in which students have rank concerns, that is, they derive direct utility from rank in terms of achievement, can explain the heterogeneous peer effects across students and subjects in a simple and intuitive way.[7] This is useful, because the patterns uncovered by the econometric model are hard to rationalize with existing models of peer effects. Moreover, this exercise demonstrates that flexibly estimated peer effects can inform theories on the mechanisms behind them.

The paper is organised as follows: section 2 presents the data and background, and section 3 describes the sample and the relationship between the earthquake and student test scores. Section 4 presents the econometric model. Section 5 explains identification, and sections 6 and 7 present the empirical findings, followed by a discussion of model fit

[6] Under the interpretation of a student's type as cost of effort (or, with the appropriate change of sign, ability), this result helps reconcile average treatment effect estimates of different signs in the literature. For example, using randomised allocation of cadets to companies in a military academy, Lyle (2009) finds that peer ability heterogeneity has, on average, positive effects on achievement. On the other hand, by manipulating the ability composition of tutorial groups among undergraduate students, Booij, Leuven, and Oosterbeek (2016) find that, on average, the standard deviation of peer ability has a negative effect on achievement. Both papers report heterogeneity of these effects. Other studies exploring the impact of the heterogeneity of peer ability include Ding and Lehrer (2007), Vigdor and Nechyba (2007), Hoxby and Weingarth (2005).

[7] See Mierendorff and Tincani (2018) for an analysis of the implications of rank concerns for ability peer effects. They test a theoretical model of rank concerns in education against experimental data and use an estimated structural model to quantify the role of rank concerns in generating peer effects.

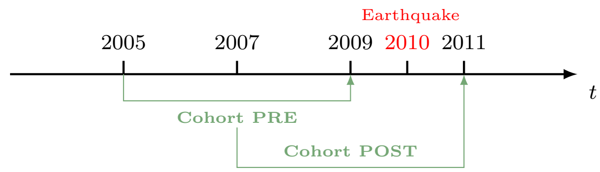


Figure 1: Data time-line.

in section 8. Section 9 discusses how the econometric model relates to existing nonlinear difference in difference models. Section 10 discusses candidate mechanisms behind the estimated impacts. Section 11 concludes.

# 2 Data and Background

## 2.1 Data

I use two cohorts of students from the SIMCE dataset (*Sistema de Medición de la Calidad de la Educación*), for a total sample size of 385, 294 students in 15, 202 classrooms. For both cohorts I observe administrative records on $8^{th}$ grade students' Mathematics and Spanish standardized test scores, father's and mother's education, household income, gender, town of residence, and lagged ($4^{th}$ grade) Mathematics and Spanish standardized test scores. Classroom level information includes class size and characteristics of the Spanish and Mathematics teachers, specifically, experience, education, tenure at the school, gender, and type of contract (permanent or probationary). School level information includes rurality, public or private status and town.

I refer to the two cohorts as pre- and post-earthquake cohorts. One cohort is observed in the $8^{th}$ grade in 2009, before the 2010 earthquake, while the other cohort is observed in the $8^{th}$ grade in 2011, after the earthquake, as shown in Figure 1. For both cohorts, identifiers are available at the student, teacher, classroom and school level, allowing me to match students to classrooms, classmates, Mathematics and Spanish teachers, and schools. Therefore, the dataset can be used to study spillovers at classroom level.

## 2.2 Earthquake

On February $27^{th}$ 2010, at 3.34 am local time, Chile was struck by a magnitude 8.8 earthquake, the fifth-largest ever instrumentally recorded and technically referred to as a mega-earthquake (Astroza, Ruiz, and Astroza 2012). Shaking was felt strongly throughout

500 km along the country, covering six regions that together make up about 80 percent of the country's population. While the death toll, as tragic as it was, was limited for such a strong earthquake (525 deaths), damage was widespread; 370,000 housing units were damaged or destroyed. The Government implemented a national reconstruction plan to rebuild or repair 220,000 units of low- and middle-income housing. Estimated total costs are around $2.5 billion. The mega-earthquake had a continued impact on people's lives during the period covered by my sample. By the time the 2011 SIMCE sample was collected, i.e., 20-22 months after the earthquake struck, despite impressive efforts by the Government, only 24 percent of home reconstructions and repairs had been completed (Comerio 2013). This led to frustration in the population, as shown in Figure 11 in Appendix B.

## 2.3 Measure of Local Earthquake Intensity

I construct a measure of the intensity of shaking in each town in the sample using the Medvedev Sponheuer Karnik (MSK) scale. An advantage of this scale is that it can be mapped into a tangible measure of disruption: the average level of damage to buildings by earthquake resistance type in each town. Because reconstruction expenses were covered by the Government, this measure of damage reflects disruptions rather than shocks to household expenses.

For a given intensity of shaking, the level of damage depends on the construction type. For example, unreinforced masonries are less resistant than reinforced masonries, therefore, the same value of MSK-Intensity corresponds to larger damages in unreinforced masonries than in reinforced ones. The type of construction of students' homes is not directly observed in my dataset. However, Astroza, Ruiz, and Astroza (2012) report that 60% of the poorest Chileans live in one of two house types with similar earthquake resistance: old traditional adobe constructions (6.1%) and unreinforced masonry houses (51.9%). Given the striking school stratification in Chile, public school students belong to the poorest 50% of Chilean households. Therefore, it is reasonable to expect that all public school students live in one of these two building types. To account for unobserved construction type, the empirical analysis restricts the sample to municipal (public) school students.[8] This sample restriction addresses potential measurement error deriving from the unobservability of students' home types. Non-random location choices of parents are

[8] As a measure of damages to individual buildings, MSK-Intensity may still contain residual noise because it averages damages across buildings within a town and, therefore, it may be vulnerable to classical measurement error inducing attenuation bias.

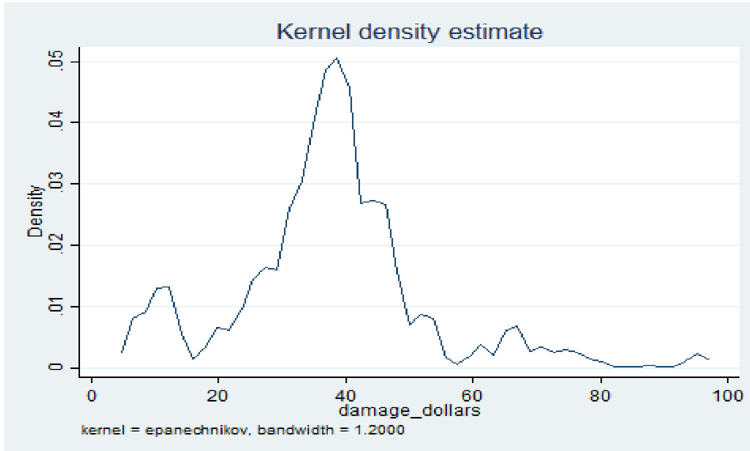


Figure 2: *Source*: SIMCE dataset and author's calculations. The right tail is truncated at USD 100 ($\sim 85^{th}$ percentile of the untruncated distribution) for ease of exposition.

a remaining but separate concern that the empirical model is specifically designed to address.

To construct MSK-Intensity, I apply the intensity attenuation formula for the Chilean 2010 earthquake, which is a function of a town's distance from the earthquake's asperity. Using the geographic coordinates of each town and of the asperity, I compute MSK-Intensity $I$ according to: $I = 19.781 - 5.927 \log_{10}(\Delta_A) + 0.00087\Delta_A$ ($R^2 = 0.9894$), where $\Delta_A$ is the distance from the main asperity. The formula is valid only at the town level and only for towns in the six regions affected by the earthquake (Astroza, Ruiz, and Astroza 2012). There are two advantages to using this measure of shaking intensity as opposed to simple distance from the asperity. First, shaking intensity is a non-linear function of distance, therefore, using distance would introduce a non-classical measurement error. Second, the MSK-Intensity measure, coupled with other formulae borrowed from the structural engineering literature, allows me to express shaking in terms of the dollar amount of damage, which has an intuitive interpretation.

## 2.4 Distribution of Earthquake Damages

Figure 2 shows the distribution of earthquake intensities among the students in my sample, which, for illustrative purposes, are expressed in terms of reconstruction expenses in US dollars. Intensities in the towns of the schools are also available and used in the analysis. On average, damages to homes are large, USD 170, equivalent to 24 percent of average household monthly income. The damage distribution is right-skewed, with a

median of USD 39, 6 percent of income, and a $90^{th}$ percentile of USD 303, 43 percent of income.
$3,822$ out of the $5,574$ classrooms in earthquake regions have a geographically dispersed student body. In those classrooms, not all classmates reside in the same town, and this generates variation in the measure of MSK-Intensity within the classroom. I use this variation to identify the effect of damage dispersion in the classroom on achievement. Figure 3 shows three examples of classrooms with students who do not all reside in the same town. Because of differences in soil type across towns, even classmates who live close to each other suffered different levels of damage. For example, the bottom panel of the Figure shows that students of the *La Florida* school who live 5.2 km apart from each other suffered a damage difference of USD 272, or 39 percent of average income. Large differences among neighbouring towns are not unusual, especially in areas closer to the asperity.[9] Among classrooms with a geographically dispersed student body, the within classroom standard deviation in damages is, on average, USD 79. Figure B in Appendix B shows the location of these classrooms on a map. The map displays three important features: first, there are many data points (represented by dots) on the map, i.e. many schools have a geographically dispersed student body; second, dots have different sizes, i.e. there is across school variation in within school damage dispersion; third, even schools close to each other suffered different damage dispersions. These are all features that will be exploited in identification.

## 2.5 Sample Restriction: No Earthquake-induced Displacements

I obtained from the Ministry of Education the list of the schools that closed as a consequence of the earthquake, as well as the list of students at those schools.[10] I observe in what schools the evacuated students enroll, and drop both the collapsed and receiving schools from the sample, for a total of 803 dropped schools, corresponding to 12 percent of the sample. This ensures that in my sample there are no earthquake-induced relocations of evacuated students across schools.[11] Such relocations could have large direct impacts on

[9] Damage dispersion effects are calculated controlling for damage in the school town and average damages in the classroom, to account for correlation between these two variables and dispersion in damages.

[10] They closed either because the buildings became unsafe, or because most of the students' homes were so badly damaged, that students had to relocate, reducing attendance below the operational minimum.

[11] These schools are dropped from both the pre- and the post-earthquake cohorts. Imberman, Kugler, and Sacerdote (2012) use the influx of Katrina evacuees in a school as an exogenous source of change to classroom composition. In Chile evacuees were spread across such a large number of schools that the

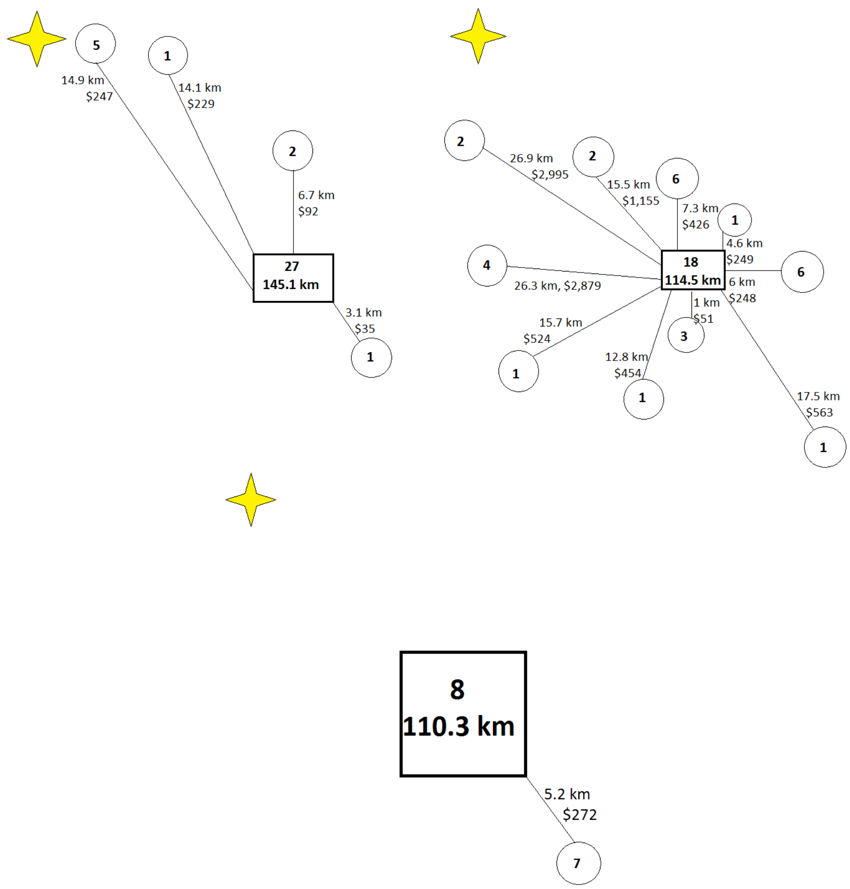


Figure 3: Examples of three schools where not all students are residents in the school town. At the top left is Colegio Santa Ines in the town of San Vicente, at the top right is Liceo Maria Auxiliadora in Santa Cruz, and at the bottom is Escuela La Florida in Talca. The squares represents the school town location relative to the earthquake asperity (star). In the square there are the number of classmates residing in the school town, and its distance from the asperity. The circles represent towns of residence of $x$ classmates, where $x$ is the number in the circle. The lines indicate the distance to school of each town, and the difference in damages suffered by an unreinforced masonry construction.

the evacuated students, and they could have spillover impacts on the incumbent students in receiving schools.

# 3 Data Description

Table 1 presents country-wide descriptive statistics for the samples of public school students and teachers in the pre- and in the post-earthquake cohorts. Student characteristics are fairly similar across cohorts. One of the main differences is a 5.8% increase in aver-

influxes in each school are too small to detect any statistically significant impact.

age income, which is in line with the countrywide +5.5% increase in GDP per capita for the same period (2009-2011). Another difference is a decrease in lagged test scores, which may indicate movement of students across school sectors. Teacher characteristics are fairly stable across cohorts, except for an increase in the proportion of teachers with post-graduate degrees.

Table 1: Summary statistics, public schools, all regions

| | Pre-earthquake cohort | | Post-earthquake cohort | |
|---|---|---|---|---|
| | mean | sd | mean | sd |
| Math Test Score | -0.32 | 0.91 | -0.32 | 0.90 |
| Spanish Test Score | -0.25 | 0.95 | -0.25 | 0.95 |
| Lagged Math Test Score | -0.08 | 0.93 | -0.13 | 0.92 |
| Lagged Spanish Test Score | -0.08 | 0.92 | -0.10 | 0.92 |
| Father's Education (years) | 9.68 | 3.42 | 9.65 | 3.15 |
| Mother's Education (years) | 9.64 | 3.27 | 9.63 | 3.12 |
| Monthly Household Income (USD) | 348.49 | 336.94 | 367.33 | 347.11 |
| % Female Math Teachers | 0.58 | 0.49 | 0.56 | 0.50 |
| % Postgraduate Degree Math Teachers | 0.53 | 0.50 | 0.64 | 0.48 |
| Teaching Experience (years) Math Teachers | 22.48 | 12.96 | 22.16 | 13.77 |
| Tenure at school (years) Math Teachers | 12.17 | 11.24 | 12.17 | 11.48 |
| % Female Spanish Teachers | 0.82 | 0.38 | 0.83 | 0.37 |
| % Postgraduate Degree Spanish Teachers | 0.51 | 0.50 | 0.59 | 0.49 |
| Teaching Experience (years) Spanish Teachers | 21.48 | 13.32 | 21.33 | 13.55 |
| Tenure at school (years) Spanish Teachers | 11.95 | 11.22 | 12.01 | 11.25 |

## 3.1 Earthquake and Test Scores

Preliminary data analysis suggest two patterns. First, earthquake exposure appears to be detrimental to test scores. Figure 4 visualizes results from a difference-in-difference regression model that indicate that the counterfactual test score growth between $4^{th}$ and $8^{th}$ grade would have been higher in earthquake regions had there not been an earthquake. Specifically, being exposed to the earthquake is estimated to have decreased test score growth by 0.026 (p-value 0.004) standard deviations in Spanish and by 0.037 (p-value 0.000) in Mathematics. Moreover, additional regressions that exploit variation in earthquake intensity within earthquake regions indicate that every 100 USD in damages to a student's home are associated with a reduction of 0.02 standard deviations in test scores ($p < 0.05$ for Mathematics, $p > 0.10$ for Spanish).[12] These findings are compatible

[12] Full regression Tables are available upon request.

with results from the medical literature, that indicate that earthquake survivors suffer from Post Traumatic Stress Disorder (PTSD) which may last for several months,[13] and that PTSD is more severe for individuals who live closer to the epicenter (Groome and Soureti 2004).

**Test score growth between $4^{th}$ and $8^{th}$ grade.**

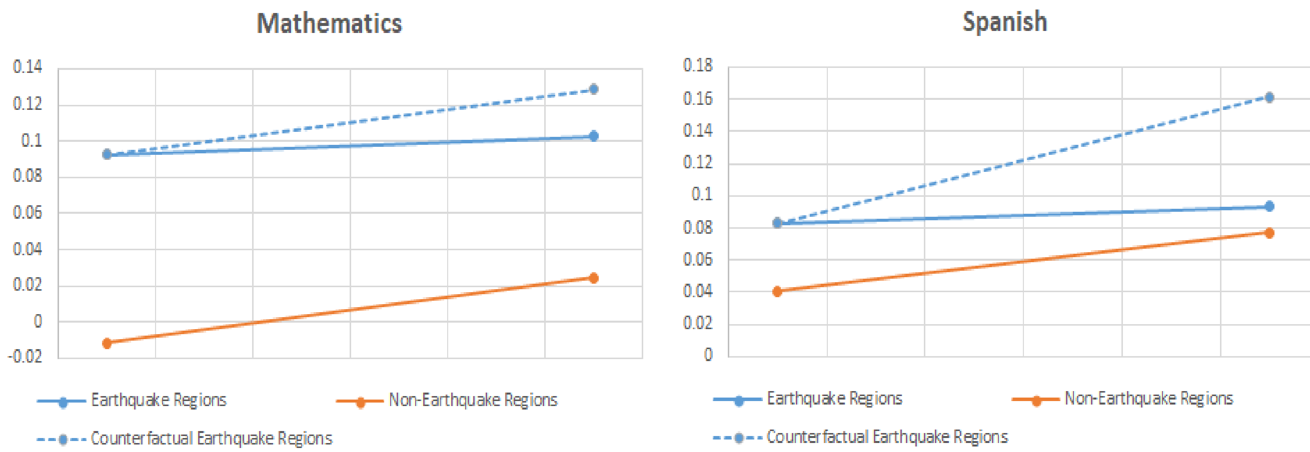


Figure 4: The y-axis shows average growth between $4^{th}$ and $8^{th}$ grade standardized test scores, while the x-axis and the colors define sub-samples over which these trends are calculated. There are only two data points on the x-axis: pre-earthquake cohort (left-most point) and post-earthquake cohort (right-most point). The lines connect them for visual clarity only. The sample is restricted to public school students. Under the identifying assumption that differences in test score trends across cohorts do not vary by region type (earthquake and non-earthquake), a counterfactual test score trend can be calculated for earthquake regions in the absence of an earthquake. The dashed line indicates that test score trends would have been higher in earthquake regions in the absence of the earthquake.

The second pattern indicates that, keeping fixed own exposure to the earthquake, the exposure of classmates may matter for own achievement. Because everyone in a classroom is either exposed or not exposed, the variation needed to identify spillovers must come from variation in intensity among classmates. Specifically, I examine two peer variables: the average and the dispersion in intensity of exposure. Using a difference-in-difference-in-differences model with continuous treatment that exploits variation across grades, cohorts, and geographic locations within earthquake regions, I obtain the estimated impacts reported in Figure 5.[14] This preliminary analysis suggests that, first, the impact of average exposure of classmates on own test scores is a precisely estimated zero. Second, dispersion in earthquake exposure among classmates appears to be detrimental to own learning,

[13] See, for example, Altindag, Ozen, et al. (2005) Lui, Huang, Chen, Tang, Zhang, Li, Li, Kuang, Chan, Mechelli, et al. (2009), Giannopoulou, Strouthos, Smith, Dikaiakou, Galanopoulou, and Yule (2006).

[14] Regression Tables and details of the estimation models are available upon request.

but these estimates are marginally significant or insignificant depending on model and outcome (Spanish or Mathematics). To interpret the magnitudes of the point estimates reported in Figure 5, back-of-the-envelope calculations indicate that moving up one standard deviation in the distribution of dispersion in model 4 decreases test scores by 0.01 to 0.08 standard deviations (depending on specification).

**Impact on test scores of the mean and of various measures of dispersion of peer earthquake exposure**

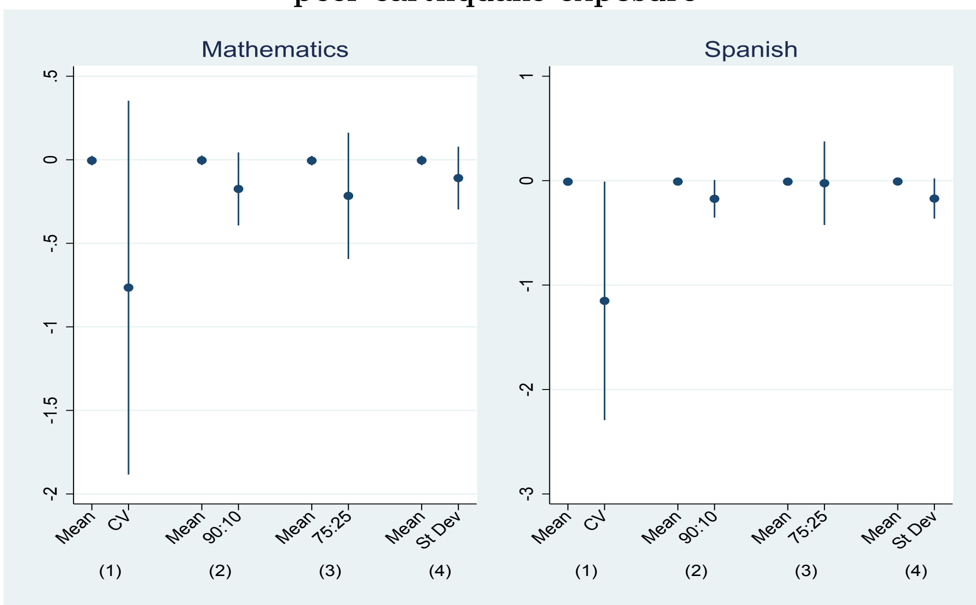


Figure 5: Estimated impacts of the mean and of the dispersion in peer earthquake exposure from four models that use different measures of dispersion: coefficient of variation (1), range between the $90^{th}$ and the $10^{th}$ percentile (2), interquartile range (3) and standard deviation (4). All regressions include controls for individual lagged test scores and characteristics, teacher characteristics and the distribution of classmate characteristics, including variables that may be correlated with dispersion in earthquake intensity like income variance. Standard errors are clustered at the school level. 95 percent confidence intervals reported.

# 4 Econometric Model

First, I define treatment as an increase in earthquake damage dispersion in the classroom and show how to identify and estimate treatment effects as a nonparametric function of student types. A student's type is discussed in detail in section 4.4. In brief, a type is a single index of student characteristics. Second, I show that earthquake damage dispersion is estimated to shift student type dispersion, so that the effect of damage dispersion can be interpreted as a contextual/exogenous peer effect in the terminology of Manski (1993), that is, the effect of dispersion in student types.

This method can be used to estimate the impact of any moment of the distribution of student types. Simply substitute the variance of MSK-Intensity with any moment of the MSK-Intensity distribution in the explanation below, and, post-estimation, verify that the chosen moment in the MSK-Intensity distribution is a shifter to the corresponding moment in the student type distribution.

## 4.1 Definition of Treatment and Outcome Equations

Define treatment as an increase in damage dispersion in the classroom. Consider a pair of classrooms, $\{r, r'\}$, with damage variances equal to $\sigma_H^2$ and $\sigma_L^2$, and with $\sigma_H^2 - \sigma_L^2 = \delta > 0$. $\delta$ is the intensity of treatment. Damage dispersion is measured by the classroom variance in MSK-Intensity. Let the indicator $G_{r,r'}$ be equal to 1 if classroom $r$ has the higher MSK-Intensity variance within the pair $\{r, r'\}$, i.e., $\sigma_r^2 = \sigma_H^2$, and 0 otherwise, and let $T_r = 1$ if classroom $r$ is in the post-earthquake cohort and $T_r = 0$ if it is in the pre-earthquake cohort. Students in the pre-earthquake cohort have not been affected by the earthquake: variance in MSK-Intensity reflects geographic dispersion, because it is calculated from students' home locations, but it does not reflect damage dispersion. All students in this cohort are untreated, therefore, they serve as the control group. The observed outcome (achievement) of student $i$ in classroom $r$ is $Y_{ir}$. Let $Y_{ir}^1$ denote the potential outcome of student $i$ if she is treated, and $Y_{ir}^0$ the potential outcome if she is not treated.

The observed outcome is:

$$Y_{ir} = \tau_{rr'} Y_{ir}^1 + (1 - \tau_{rr'}) Y_{ir}^0, \tag{1}$$

where the treatment indicator is $\tau_{rr'} = T_r \times G_{r,r'}$: student $i$ is treated if she is in the classroom with the higher MSK-Intensity variance in the post-earthquake cohort, untreated otherwise.

I assume that potential outcomes satisfy:

$$Y_{ir}^0 = h^0(c_i, W_r, T_r, \sigma_r^2) + \epsilon_{ir}^0 \tag{2}$$

$$Y_{ir}^1 = h^0(c_i, W_r, T_r, \sigma_r^2) + \lambda^{DD}(c_i, W_r, \delta) + \epsilon_{ir}^1 \tag{3}$$

where $c_i$ is a student's type, that is, a scalar student characteristic and $W_r$ is a vector of classroom characteristics, including peer characteristics. No distributional assumptions are made on the error terms, which are only assumed to be mean-independent:

$E[\epsilon_{ir}^0|c_i, W_r, T_r, \sigma_r^2] = E[\epsilon_{ir}^1|c_i, W_r, T_r, \sigma_r^2] = 0$.[15] From equations (2) and (3), it is clear that the mean treatment effect is given by the function $\lambda^{DD}(c_i, W_r, \delta)$ (where $DD$ stands for damage dispersion). The treatment effect depends on student and classroom characteristics, and on treatment intensity. The objective of this analysis is to describe how the treatment effect varies with student type $c_i$. Therefore, the estimator will average out $W_r$ and $\delta$. However, with a large enough number of classrooms, this econometric framework can be used to non-parametrically estimate how treatment varies with classroom characteristics and/or with treatment intensity.

I assume the following structure for the potential outcome function $h^0(\cdot)$:

$$\begin{aligned} h^0(c_i, W_r, T_r, \sigma_r^2) &= \lambda(c_i, W_r, \sigma_L^2) + T_r \cdot \left[\lambda^E(c_i, W_r) + \lambda^{DD}(c_i, W_r, \sigma_L^2)\right] + \\ &\quad + G_{r,r'} \cdot \lambda(c_i, W_r, \delta). \end{aligned} \tag{4}$$

The model in equations (2), (3) and (4) makes minimal assumptions on how student and classroom characteristics combine to produce outcomes and on their interaction with the treatment variable. This is what allows the model to identify any pattern of heterogeneity of treatment effects in the classroom. Specifically, the model embeds two assumptions on the production of achievement:

*Assumption 1.* $h^0(\cdot)$ is additively separable in $\lambda(\cdot), \lambda^E(\cdot)$ and $\lambda^{DD}(\cdot)$.
*Assumption 2.* $\lambda(\cdot, \cdot, \sigma^2)$ and $\lambda^{DD}(\cdot, \cdot, \sigma^2)$ are linear in $\sigma^2$.

Additive separability (*Assumption 1*) is what permits identification of the treatment effect function through double differences. *Assumption 2* is what allows me to express the realised outcome as:

$$\begin{aligned} Y_{ir} = h_r(c_i) &= \lambda(c_i, W_r, \sigma_L^2) + T_r \cdot \left[\lambda^E(c_i, W_r) + \lambda^{DD}(c_i, W_r, \sigma_L^2)\right] + \\ &\quad + G_{r,r'} \cdot \lambda(c_i, W_r, \delta) + T_r \cdot G_{rr'} \cdot \lambda^{DD}(c_i, W_r, \delta) + \epsilon_{ir}. \end{aligned} \tag{5}$$

Recognize that equation (5) has a similar structure to a typical linear difference in differences model (D-i-D): there is a cohort dummy ($T_r$), a dummy that indicates the low or high variance status ($G_{rr'}$), and an interaction of these two dummies. The cohort represents the group (treatment or control), while the variance status dummy plays the role of time in a typical linear D-i-D model. Thanks to *Assumption 2*, we can separate the

[15] This is a weaker assumption than the full-independence assumptions imposed in the nonlinear difference-in-differences models in Athey and Imbens (2006).

functions $\lambda(c_i, W_r, \cdot)$ and $\lambda^{DD}(c_i, W_r, \cdot)$ into two additive components, one that multiplies 1 and $G_{rr'}$ for $\lambda$, and one that multiplies $T_t$ and $T_r \cdot G_{rr'}$ for $\lambda^{DD}$.

The main differences with typical D-i-D models are that functions replace the regression coefficients, cross-sectional comparisons replace time trends, and the treatment indicator $\tau_{rr'} = T_r \times G_{rr'}$ has double index $rr'$ because treatment status is determined within each pair of classrooms. The estimand of interest is the function $\lambda^{DD}(c_i, W_r, \delta)$, which multiplies the treatment indicator $T_r \cdot G_{rr'}$.

The functions on the right had side of (5) have the following interpretations. $\lambda(c_i, W_r, \sigma^2_L)$ is the counterpart of the constant in a linear D-i-D: it is the outcome function for students in the control group (pre-earthquake cohort) who are not subject to treatment, that is, who are in the lower-variance classroom within the pair $rr'$. In these pre-earthquake classrooms, $\sigma^2_L$ reflects the geographic dispersion of the student body.

When this function is evaluated in the classroom with the higher MSK-Intensity variance within the pair, $\lambda(c_i, W_r, \sigma^2_H = \sigma^2_L + \delta)$, it traces the outcomes of pre-earthquake cohort students ($T_r = 0$) who are in the higher-variance classroom within the pair ($G_{rr'} = 1$). The additional geographic dispersion effect (GDE) experienced by the higher-variance classroom compared to the lower-variance classroom, $\lambda(c_i, W_r, \delta)$, is identified from the difference in outcomes between students in higher- and lower-variance classrooms in the pre-earthquake cohort. In equation (5), this is the GDE function that multiplies the dummy $G_{r,r'}$. GDE are the counterpart of time trends in typical D-i-D models.

$\lambda^E(c_i, W_r)$ are cohort effects, the counterpart of group effects in a typical linear D-i-D. They let the achievement of post-earthquake cohort students differ from the achievement of pre-earthquake cohort students for any number of reasons. An important example is that the earthquake could have had a direct effect on student achievement. No restrictions are made on how these effects vary by student and classroom characteristic. For example, high $c_i$ students may have been affected more (or less) by the earthquake than low $c_i$ students; more affluent schools may (or may not) have attenuated the impact of the earthquake; teachers may (or may not) have focussed their attention on the most vulnerable students, and so on. Another source of cohort effects are policy changes occurring in between the two cohorts, like the amendments to the voucher system implemented in 2008. These are only two examples of possible sources of cohort effects. The treatment effects are robust to any change occurring between cohorts. Moreover, the functional form of these cohort effects is entirely unrestricted, allowing the researcher to be agnostic about how they vary across classrooms and students.

Finally, $\lambda^{DD}(c_i, W_r, \delta)$ is the treatment effect function calculated at treatment intensity

$\delta$. It measures the impact of an increase in damage dispersion by $\delta$ on a student of characteristics $c_i$ who is in a classroom with characteristics $W_r$. The treatment effect function measures the causal impact of moving from a classroom with damage dispersion $\sigma_L^2$ (where the impact of $\sigma_L^2$ on achievement is measured by $\lambda^{DD}(c_i, W_r, \sigma_L^2)$ in equation (5)) to one with damage dispersion $\sigma_L^2 + \delta$, keeping $W_r$ constant. $W_r$ includes peer characteristics such as their lagged test scores and socio-economic status.

To summarize, in this flexible nonlinear model, estimates of treatment effects as a function of student type $c_i$ are robust to cohort effects and geographic dispersion effects which can be arbitrarily heterogenous across students and classrooms. Cohort effects include any earthquake impact on student outcomes that is not mediated by damage dispersion effects. Having a model that is robust to these kind of effects is important, because an earthquake is a complex phenomenon which could impact achievement through a number of channels.

## 4.2 Double Difference of Functions

Within each classroom $r$, the different components of $h(\cdot)$ described in equation (5) are not separately identified. However, the treatment effect function $\lambda^{DD}(\cdot)$ is identified through double differences: one *within* and one *across* cohorts. Consider two pairs of classrooms: one pair in the post-earthquake cohort, $r$ and $r^{'}$, and one pair in the pre-earthquake cohort, $s$ and $s^{'}$. Assume that $G_{rr'} = 1$ and $G_{ss'} = 1$, that is, $r$ and $s$ are the classrooms with the relatively higher variance in MSK-Intensity within their pair. Moreover, assume that $\delta_{rr'} = \delta_{ss'} = \delta$, that is, the treatment intensity within pair is the same across pairs. Finally, assume that, except for the variance in MSK-Intensity, these four classrooms share all other characteristics, that is $W_r = W_{r'} = W_s = W_{s'} = W$. In the empirical implementation the $W_r$ vector has 14 elements: mean, variance, skewness and kurtosis of parental education of peers and of lagged test scores of peers, average income in class, percentage of female students in class, average MSK-Intensity in class, class size, MSK-Intensity in the school town, teacher's teaching experience in years. Conditional on $W$ and $\delta$, the damage dispersion effect can be obtained as a function of $c_i$ through the following double difference of functions, performed point by point:

$$\begin{aligned} E\left[\lambda(c_i; W, \delta)|c_i, W, \delta\right] &= E\left[\Big(h_r(c_i) - h_{r'}(c_i)\Big) - \Big(h_s(c_i) - h_{s'}(c_i)\Big)|c_i, W, \delta\right] \\ &= \Big(\lambda^{GD}(c_i; W, \delta) + \lambda(c_i; W, \delta)\Big) - \lambda^{GD}(c_i; W, \delta) \end{aligned} \tag{6}$$

where the expectation is taken with respect to the $\epsilon_{ir}$ shocks. The double difference is visualised in two Figures: Figure 6 shows the *within* cohort differences that identify the confounding GDE from the pre-earthquake classrooms, and Figure 7 shows the *across* cohort differences that nett out the GDE. These differences are taken for each value of $c_i$, therefore, they require matching students *within* and *across* cohorts based on their student type $c_i$.

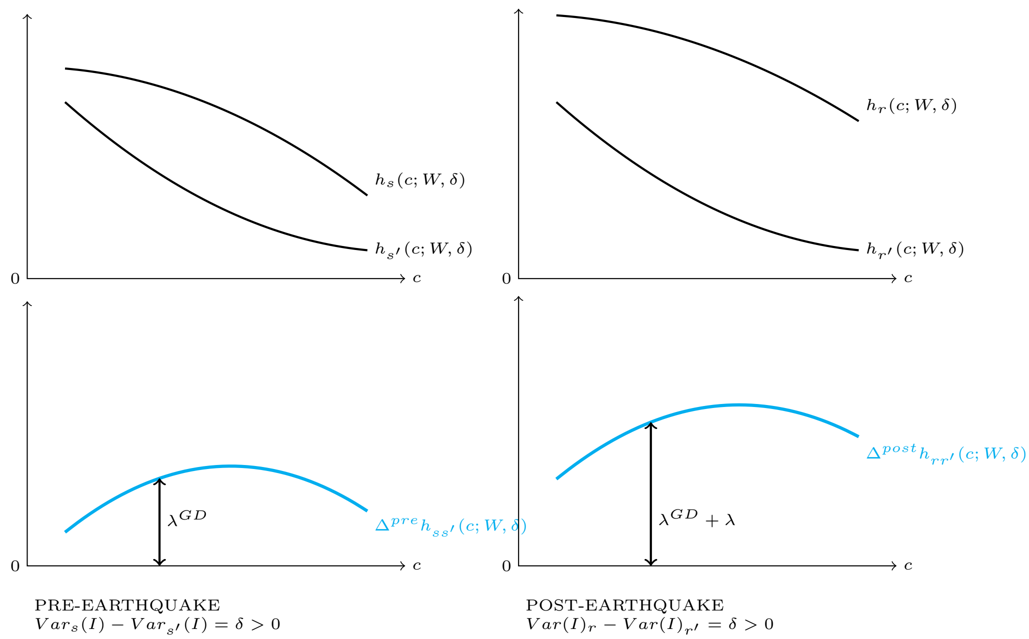


Figure 6: Classrooms $r, r', s$, and $s'$ have identical $W$ and within-pair $\delta$. $\lambda^{GD}$ is the geographic dispersion effect function, $\lambda$ is the effect on achievement of increasing damage dispersion by $\delta$, as a function of student type $c$, conditional on $W$.

## 4.3 Obtaining Treatment Effects as a Function of Student Type

For each pair of pairs of classrooms in the data that are matched on $W$ and $\delta$, the effect of damage dispersion on achievement as a function of student's type $c_i$ is identified through the conditional double difference in 6. Matching on $W$ and $\delta$ addresses any unbalance in the distribution of covariates between the control (pre-earthquake) and the treatment (post-earthquake) groups (Smith and Todd 2005). These covariates are then integrated out using their empirical distribution. Let $f(W, \delta)$ indicate the empirical distribution of quadruplets of classrooms with the same $W$ and $\delta$. The effect of increasing damage dispersion, unconditional on $W$ and $\delta$, is obtained by averaging the treatment effect over

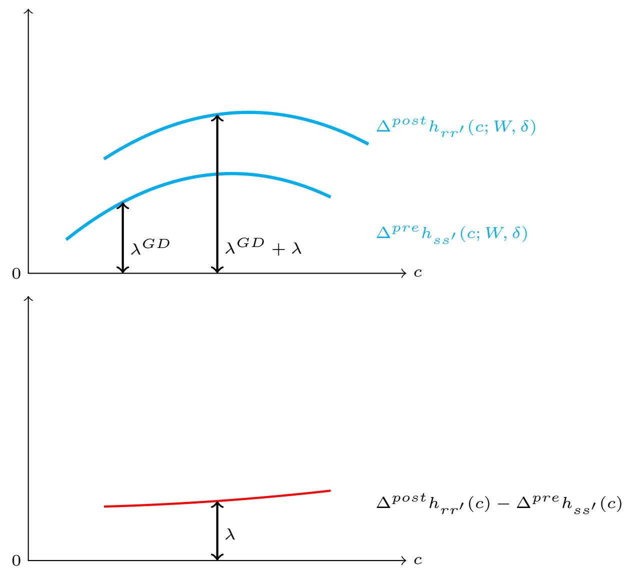


Figure 7: Netting out the geographic dispersion effects. Notice that the difference between the $\Delta h$ functions can be taken only over the overlapping portion of the two domains. The nonlinear difference-in-differences models in Athey and Imbens (2006) impose a similar support restriction.

$W$ and $\delta$, with $\delta > 0$:

$$TE(c_i) = E[\lambda(c_i; W, \delta)] = \int \lambda(c_i; W, \delta) I[\delta > 0] f(W, \delta) dW d\delta \tag{7}$$

where $TE(c_i)$ is the treatment effect, and $I[\cdot]$ is an indicator function equal to 1 if its argument is true. In practice, matching quadruplets of classrooms with respect to $W$ and $\delta$ is performed by kernel weighting, in the spirit of Ahn and Powell (1993). Because of the high dimensionality of $W$, it would be difficult to find a quadruplet of classrooms that are exactly identical in all elements of $W$; for this reason, I use nearest neighbour matching.[16] Weights are built with multivariate standard normal kernel functions. Details of the weighting procedure can be found in Appendix A.3.

For now, assume that nonparametric estimates of the $h_r(\cdot)$ functions are available for all classrooms in the sample. Given kernel weights $\omega_{rr'ss'}$ for each quadruplet of classrooms in the sample (two from each cohort), at each candidate value of $c$ the estimate of the

[16]One additional regularity condition is required to apply kernel matching: the function $\lambda^E(c_i, W_r)$ must be continuous in $W_r$, and the functions $\lambda^{GD}(c_i, W_r, \sigma_r^2)$ and $\lambda(c_i, W_r, \sigma_r^2)$ must be continuous in $W_r$ and $\sigma_r^2$. In Ahn and Powell (1993), this assumption corresponds to the continuity of the selection function (see page 9 of their paper). This guarantees that there are no jumps when we compare pairs of classrooms that are similar but not identical in $W$ and $\delta$.

treatment effect $TE(c)$ is obtained through the following sample mean:

$$\hat{TE}(c) = \frac{\sum_{s=1}^{N^{pre}-1} \sum_{s'=s+1}^{N^{pre}} \sum_{r=1}^{N^{post}-1} \sum_{r'=r+1}^{N^{post}} \omega_{rr'ss'}(\Delta^{post}\hat{h}_{rr'}(c;W,\delta) - \Delta^{pre}\hat{h}_{ss'}(c;W,\delta))}{\sum_{s=1}^{N^{pre}-1} \sum_{s'=s+1}^{N^{pre}} \sum_{r=1}^{N^{post}-1} \sum_{r'=r+1}^{N^{post}} \omega_{rr'ss'}} \tag{8}$$

where $N^{post}$ and $N^{pre}$ are the sample number of classrooms in the post- and pre-earthquake cohorts, and the numerator contains the double difference at point $c$.

## 4.4 Student Types and Nonparametric Estimation of the Outcome Functions

If student type $c_i$ were observed, the $h_r(c_i)$ functions in equation (6) could be estimated non-parametrically at a grid of values for $c_i$ and these nonparametric estimates could be used to estimate the treatment effect function using equation (8). To be useful for the estimation of treatment effects as a function of student type, the latter must be a variable $c_i$ that identifies comparable students both within and, importantly, *across* cohorts. This is because taking the double difference of functions point by point requires matching students on type $c_i$ within and across cohorts.

There is a difference between students across cohorts: those in the post-earthquake cohort have been affected by the earthquake, while those in the pre-earthquake cohort have not. Section 3.1 showed that individual exposure to the earthquake impacted student's achievement, therefore, it is desirable to allow local earthquake intensity experienced by a student to affect her type. However, in the data there is no variable that captures this, because, for both cohorts, all baseline student level variables were measured before the earthquake struck.

To see why this is problematic when taking *across cohort* differences, consider using lagged test score $y_{i,t-1}$ as a measure of $c_i$. It was measured before the 2010 earthquake (in 2005 for the pre-earthquake cohort and in 2009 for the post-earthquake cohort). A student in the post-earthquake cohort whose house has been badly damaged in the earthquake may have a very different productivity from a student in the pre-earthquake cohorts who has the same lagged test score but who has not been affected by the earthquake. Not taking this into account would bias the estimate of the treatment effect function. Students across cohorts cannot be matched on lagged test score $y_{i,t-1}$, nor on any other student characteristic observed in the data.

To address this issue, I model student type $c_i$ as an index function of the student characteristics that are observed in the data, and I let earthquake damage at a student's home be one of the determinants of type $c_i$ for post-earthquake students. Formally:

$$c_i = c^{T_r}(X_i;\theta) = \begin{cases} \theta_1 y_{i,t-1} + \theta_2 peduc_i + \theta_3 income_i + \theta_4 female_i, & \text{if } T_r = 0 \\ c^{T_r=0}(X_i) + \theta_5 I_i + \theta_6 I_i X_i^-, & \text{if } T_r = 1 \end{cases}$$

where $y_{i,t-1}$ is lagged achievement, $peduc_i$ is the average of father's and mother's years of education, $income_i$ is total household monthly income, $female_i$ is equal to 1 if student $i$ is female, vector $X^-$ is the vector of all student characteristics except for MSK-Intensity ($I_i$), that is, $X^- = [y_{i,t-1}, peduc_i, income_i, female_i]$. Finally, vector $X_i = [X_i^-, I_i]$ includes all student's characteristics. In the post-earthquake cohort ($T_r = 1$), MSK-intensity is one of the determinants of student type. So is the interaction between MSK-intensity and student characteristics, which allows for heterogeneity in how damage affects a student's type $c_i$.[17] This way of modelling and estimating student type overcomes the issues discussed above that would arise if we used, instead, any of the student characteristics available in the data.[18]

There are two additional advantages to modelling student type as a single index. First, it reduces the curse of dimensionality compared to using a vector of student characteristics. This improves the precision of the estimator of the treatment effects (Abadie 2005, Horowitz 2010). Second, it allows me to obtain interpretable results, because the heterogeneity of the treatment effects can be graphed with respect to a single scalar. Typically, heterogeneity of peer effects is expressed with respect to a single student characteristic such as lagged test scores. The single index estimated here is more comprehensive because it incorporates more student variables.

The nonparametric function $h_r(c_i)$ combined with the parametric model for $c_i$ are a semi-parametric single index model. Under regularity conditions set out in Ichimura (1993), conditional on $W_r$, $T_r$, and $\sigma_r^2$, the mapping from $c_i$ to achievement, $h_r(c_i)$ in equation (5), and the parameters of the index $c^{T_r}(X_i;\theta)$ are identifiable for all $r$.[19] Appendix A.1 describes the algorithm for the estimation of $h_r(c_i)$. Appendix A.2 uses results on

[17] For example, wealthier parents may try to attenuate the impact of the earthquake by providing more resources to an affected child, or the psychological impact of $I_i$ may vary by gender.

[18] For example, suppose that we estimate that pre-earthquake $c_i = 1 \cdot y_{i,t-1}$ and post-earthquake $c_i = 1 \cdot y_{i,t-1} - 2 \cdot I_i$. Then, a pre-earthquake student with lagged achievement $y_{i,t-1} = 10$ is the appropriate control/match for a post-earthquake (treated) student with higher lagged achievement $y_{i,t-1} = 12$, but who is hit by an earthquake shock of 1 ($12 - 2 \cdot 1 = 10$), rather than for a post-earthquake student with the same lagged achievement $y_{i,t-1} = 10$ who is hit by the same shock ($10 - 2 \cdot 1 = 8 \neq 10$).

[19] The $\theta$ parameters are identified up to a normalisation. Here, I interpret $c_i$ as a shifter to cost of effort and set the coefficient on $y_{i,t-1}$ to $-1$, assuming that higher academic skills reduce cost of effort. The regularity conditions include assuming that $X_i$ has at least one continuously distributed component whose $\theta$ coefficient is nonzero, and $h$ is differentiable and nonconstant in $c_i$ (Ichimura 1993).

semi-parametric models to build the standard errors for the treatment effect function in equation (7). Estimating the treatment effect function $\hat{TE}(c)$ over a grid of values for $c$ allows me to trace the treatment effect $TE(c)$ as a function of student types. Computing $\hat{TE}(c)$ at each grid point requires a number of calculations of the order of $10^{12}$, therefore, parallel processing is required. Using $\sim 2,000$ nodes on the UCL Legion cluster, estimation is completed in around 70 hours.

# 5 Identification

## 5.1 Shifting the Types of Peers by Shifting their Earthquake Intensities

We would like to relate the effects of damage dispersion (or of any moment of the classroom damage distribution), a concept specific to the Chilean earthquake context, to the more broad concept of peer effects. I define peer effects as the causal effect on own achievement of changing the classroom distribution of student types $c_i$. To estimate peer effects, we need an exogenous shifter to the variance (or other desired moment) of student types. Having modelled student types as a single index, it is possible to shift its distribution by shifting the distribution of any of the variables that enter the index. For example, variation in the classroom distribution of parental income generates variation in the classroom distribution of student types. However, such variation is likely to be correlated with unobservables affecting outcomes.

In general, peer characteristics such as lagged test scores, gender and socio-economic status are known to correlate with unobserved determinants of student achievement. Much of the empirical peer effect literature in education is concerned with finding data variation that breaks this correlation (see, for example, Hoxby (2000)). In this paper I take a different approach. I estimate peer effects keeping these peer characteristics constant. To do so, I use variation in a component of student types that can be varied independently of these peer characteristics. That is, I vary the distribution of earthquake intensity shocks in a classroom (in particular, its variance), keeping all other peer characteristics constant. The innovation of this approach is that the unobserved characteristics that typically covary with these peer characteristics are kept constant, eliminating a well known source of endogeneity in the estimation of peer effects.[20]

[20]Consider this thought experiment. Two post-earthquake classrooms, A and B, are both affected by the earthquake. The composition of students in these two classrooms is identical: the same distribution of lagged test scores, SES level, etc., however, the earthquake exposure is different. In classroom A,

## 5.2 Identifying Assumption: Common Geographic Dispersion Effects

While the usual peer effect confounders are kept constant, using variation in the classroom variance of earthquake damages poses different challenges to the identification of causal effects. This classroom variable, too, may be correlated with unobservables that affect achievement, but in a way that the empirical model and data are designed to address. Specifically, the model allows for geographic dispersion effects (GDE), function $\lambda(\cdot)$ in equation (5). GDE are due to a mechanical positive relationship that exists between variance in earthquake damages (the treatment variable) and geographic dispersion of the student body. Mechanically, the more dispersed the student body is geographically, the larger the variance in earthquake damages is. Keeping observed classroom and peer characteristics constant, students in classrooms with different geographic dispersion can have different achievement levels because of unobserved classroom characteristics that correlate with geographic dispersion. An example for why this may be the case is that a classroom that attracts students from far away may be more desirable in terms of characteristics that the econometrician does not observe, but that the parents see and base their school choice on. In fact, a descriptive analysis of the pre-earthquake sample indicates that in classrooms with more geographically dispersed students, test scores are higher, keeping everything else constant (see Table 6 in Appendix B).

Not taking this into account would bias estimates of the treatment effect of damage dispersion (and, ultimately, of peer effects). The model and data account for potentially confounding GDE: the pre-earthquake cohort serves as a control group that experiences GDE but not damage dispersion effects, and it is used to nett out the GDE from the gross effects calculated on the treatment group (the post-earthquake cohort). The identifying assumption is the equivalent of the common trends assumption in typical D-i-D models. I call it “the common GDE across cohorts” assumption: GDE must not depend on cohort $T_r$. Notice that GDE are modelled flexibly: no assumptions need to be made on their sign

all students are affected equally by the earthquake: the variance of student types in this classroom is entirely determined by student composition. In classroom B, on the other hand, students were affected differently by the earthquake: the positive variance in earthquake exposure generates additional variance in students types, beyond what is determined by student composition alone. As a result, compared to classroom A, in classroom B there are more relatively high and relatively low type students, and fewer middle-type students, even though the classroom composition is identical to that in classroom A. By matching classrooms that have identical student compositions but different variances of earthquake intensities, it is possible to measure how the classroom achievement distribution varies with the classroom variance in student types, keeping student composition in terms of lagged test scores, gender and family background constant.

or shape. Therefore, estimates of the treatment effects are robust to GDE of any sign, and to GDE that vary arbitrarily across students and classroom characteristics. This is more general than the way time trends are typically modelled in a linear D-i-D model, where they are assumed to be constants.

## 5.3 Test of the Identifying Assumption

As usual, testing an identifying assumption is possible if additional data is available that is not used in estimation. In this context, I must be able to estimate for both cohorts the effect of geographic dispersion, in the absence of damage dispersion. This is not possible with the estimation sample, because geographic dispersion and damage dispersion co-move in the post-earthquake cohort. Instead, I use data from the regions which were never affected by the earthquake, because in this sample there are no earthquake damages and, therefore, variation in geographic dispersion is not accompanied by variation in damage dispersion. An issue that arises is that in these regions the treatment variable (variance of MSK-Intensity) cannot be calculated, because the structural engineering formula on which it is based does not exist for these regions (Astroza, Ruiz, and Astroza 2012). To overcome this issue, I must use a different measure of geographic dispersion that is available in non-earthquake regions.

I define a classroom as geographically homogeneous if all students reside in the same town, and geographically dispersed if at least one student comes from a different town. In the entire sample, 55 percent of schools are geographically dispersed according to this measure. Letting $G_r$ indicate this discrete geographic dispersion measure, I estimate the following regression model:

$$\begin{aligned} y_{ir} &= \gamma y_{ir,g-1} + \theta_0 + \theta_1 T_r + \theta_2 X_{ig} + \theta_3 X_{ig} T_r + \theta_4 Z_{rg} + \theta_5 Z_{rg} T_r + \\ &\quad \theta_6 G_r + \theta_7 G_r \times T_r + \zeta_{irg}, \end{aligned} \tag{9}$$

where $y_{ir,g-1}$ is lagged (grade 4) test score of student $i$ in classroom $r$, $y_{ir}$ is test score in grade 8, $X_{ig}$ are student characteristics, $Z_{rg}$ are classroom characteristics, and $T_r$ is equal to 1 for post-earthquake students and to 0 for pre-earthquake students. Parameter $\theta_1$ measures cohort effects, and parameters $\theta_3$ and $\theta_4$ measure cohort effects on the achievement production coefficients. They are added to make the model comparable to the semiparametric model, which allows for unrestricted changes to the achievement production functions across cohorts. The parameter of interest is $\theta_7$. When equation (9) is estimated in non-earthquake regions, parameter $\theta_7$ tests the hypothesis that average (across students) geographic dispersion effects are constant across cohorts, a necessary condition for identification.[21]

The top panel of Table 2 shows results from the estimation of equation (9) in non-earthquake regions. The third row in the Table shows that the estimate of $\theta_7$ is always small and always statistically insignificant in models with covariates (it is only significant in the model without covariates in column 3). This indicates that when covariates are included in the model, as they are in the main estimation model, GDE are, on average, constant across cohorts. This satisfies a necessary condition for identification.

As an additional check, the bottom panel of Table 2 shows results from the estimation of model (9) in earthquake regions. The goal of this check is to verify that the lack of a change in GDE across cohorts in non-earthquake regions is not due to a failure of the variable used to measure GDE. In other words, if the measure of geographic dispersion is good, it should be able to pick up damage dispersion effects when model (9) is estimated in earthquake regions. Indeed, results show that a larger geographic dispersion according to this measure has similar negative effects on test scores as those reported in Figure 5, which base the measure of GDE on the MSK-Intensity of students.

# 6 Estimation Results: Student Types

## 6.1 Parameter Estimates of the Student Type Index

Table 3 presents the estimates of the single index parameters. The coefficient on lagged test score, $\theta_1$, has been normalized to $-1$. Parental education and income are estimated to have an impact on student type of the same sign as that of lagged test score, while earthquake intensity is estimated to have an impact of the opposite sign, as expected, but the point estimate is not significant. The interactions of earthquake intensity with income and gender are significant for Mathematics and Spanish test scores respectively, indicating that lower income students and female students are those whose type is affected more by the earthquake.[22] The finding on gender differences in the interaction is compatible with the medical literature. For example, on a sample of young adults who survived the

[21] For numerical tractability, I use a linear regression model which tests if the identifying assumption holds on average.

[22] Because of the normalization, a negative coefficient should be interpreted as "good" for a student, a positive coefficient as "bad". To bootstrap the standard errors, I must account for the clustered sample design. To do so, I bootstrap 100 samples stratified at the classroom levels, and I estimate $\theta$ in each bootstrapped sampled.

Table 2: Testing the identifying assumption using non-earthquake regions. Sample of Municipal schools.

| | (1)<br>Math TS | (2)<br>Math TS | (3)<br>Spanish TS | (4)<br>Spanish TS |
|---|---|---|---|---|
| **Non-earthquake regions** | | | | |
| $G_r \times T_i$ | -0.014 | 0.013 | -0.030$^{+}$ | -0.010 |
| | (0.017) | (0.018) | (0.018) | (0.019) |
| Lagged TS | | 0.664$^{***}$ | | 0.679$^{***}$ |
| | | (0.005) | | (0.005) |
| Observations | 47,396 | 23,473 | 47,253 | 23,298 |
| **Earthquake regions** | | | | |
| $G_r \times T_i$ | -0.088$^{***}$ | -0.034$^{**}$ | -0.094$^{***}$ | -0.017 |
| | (0.011) | (0.011) | (0.012) | (0.012) |
| Lagged TS | | 0.666$^{***}$ | | 0.684$^{***}$ |
| | | (0.003) | | (0.003) |
| Observations | 110,320 | 58,783 | 110,748 | 56,805 |
| Controls | No | Yes | No | Yes |

Standard errors in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Controls: whether the student lives in the same town of the school, mother's and father's education, household income, student gender, class size, the teaching experience and tenure at the school of (resp.) the Math and Spanish teacher, whether he/she is female, has a postgraduate degree, has a permanent contract. A constant is included.

L'Aquila 2009 earthquake, females were significantly more likely to suffer from PTSD (Dell'Osso, Carmassi, Massimetti, Daneluzzo, Di Tommaso, and Rossi 2011). Finally, the coefficient on gender is of opposite signs across subjects, indicating that females have a lower type than males in Spanish, and *vice versa* in Mathematics.

Table 3: Parameter Estimates (bootstrapped standard errors in parentheses)

| Parameter | Coefficient on | Mathematics | Spanish |
|---|---|---|---|
| $\theta_2$ | Parental Education | $-0.0116^{***}$ | $-0.0212^{***}$ |
| | | (0.0052) | (0.0045) |
| $\theta_3$ | High Income Dummy | $-0.0560^{***}$ | $-0.0356^{**}$ |
| | | (0.0162) | (0.0175) |
| $\theta_4$ | Female | $0.1290^{***}$ | $-0.2303^{***}$ |
| | | (0.0195) | (0.0350) |
| $\theta_5$ | MSK-Intensity | 0.0326 | 0.0946 |
| | | (0.0596) | (0.1438) |
| $\theta_{6_1}$ | MSK-Intensity*High Income | $-0.0004^{***}$ | -0.0004 |
| | | (0.0000) | (0.0027) |
| $\theta_{6_2}$ | MSK-Intensity*Female | -0.0031 | $0.0550^{*}$ |
| | | (0.0288) | (0.0334) |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## 6.2 Interpretation of Student Type

The analysis is agnostic about and independent from the interpretation of student type. However, for interpretation of the treatment effects, it is useful to relate the concept of type to a more concrete determinant of student achievement.

Type is decreasing in lagged test scores, parental education and parental income, its relationship with gender depends on the subject, and, according to point estimates, it increases with earthquake intensity, more so for females and low income students. Therefore, a natural interpretation is that type decreases student productivity. In standard models of achievement production, achievement is typically monotonically increasing in student productivity (see, for example, the seminal work in Arnott and Rowse (1987)).[23]

[23] For example, suppose that $c_i \leq 0$ represents productivity. Achievement is produced as $y_i = \alpha e_i c_i$, where $e_i$ is effort and where $\alpha > 0$. Productivity affects the rate at which one additional unit of effort is transformed into achievement. Students maximize utility with respect to effort. Utility is the difference between achievement and cost of effort: $u_i = y_i - \frac{1}{2}\lambda e_i^2$, with $\lambda > 0$. At the optimum, achievement is monotone increasing in student type $c_i$: $y_i^* = 2\frac{\alpha^2}{\lambda}c_i^2$. Alternatively, suppose that $c_i$ represents a student's shifter to the cost of study effort. Suppose that achievement is produced as $y_i = \beta e_i$, and students maximize $u_i = y_i - \frac{1}{2}\lambda c_i e_i^2$. In this case, optimal effort is decreasing in $c_i$: $e_i^* = \frac{\beta}{\lambda c_i}$. As a result, at the optimum, achievement is monotonically decreasing in student type $c_i$: $y_i^* = \frac{\beta^2}{\lambda c_i}$.

This means that, under this interpretation of student type, we would expect achievement to be monotonically decreasing in student types. Because the econometric model does not impose monotonicity ex-ante, I can test for it post-estimation.

Figure B in Appendix B shows visually an example of two estimated classroom-specific functions $\hat{h}_r(\hat{c}_i)$. As can be seen, the higher a student's type $\hat{c}_i$ is, the lower achievement is. I formally test monotonicity of $h_r(c_i)$ in $c_i$ using the method developed in Chetverikov (2018). The null hypothesis that the $h$ function is monotonically decreasing in $c_i$ is not rejected at the $\alpha = 0.10$ significance level (see Appendix C).

This post-estimation check fits nicely with survey evidence not used in estimation, suggesting that seismic intensity at a student's home affected a student's self-reported cost of exerting study effort. Students were asked to rate how much they agree with sentences such as "It costs me to concentrate and pay attention in class" and "Studying Mathematics costs me more than it costs my classmates." Combining the answers to these questions into a single factor, I find that conditional on student lagged test scores and parental education and income, post-earthquake students affected by a higher earthquake intensity report that it is more costly for them to study, as shown in Table 7 in Appendix B.

## 6.3 Damage Variance Shifts the Variance of Student Types

To interpret the outcome of the treatment effect estimator in (8) as a peer effect of increasing the variance in student types, it must be that damage dispersion is a shifter to the variance of student types. This is an empirical question, because it depends on the estimated $\theta$ parameters in the student type function. To answer this question, I use the estimated parameters from Table 3 to predict student types $\hat{c}_i$ $\forall i$ and to calculate the variance of student types in each post-earthquake classroom. I then regress this variance on the variance of earthquake damages, controlling for the classroom characteristics $W_r$ that are controlled for in the main empirical model.

Results from this regression are reported in Table 4. They indicate that the variance in predicted student types is increasing in the variance of earthquake damages. While the linear models reported in columns (1) and (3) are unable to capture how variance in types depends on variance in damages, columns (2) and (4) reveal that when allowed to enter quadratically, the variance in damages does indeed impact the variance in student types. This function is increasing and concave over the support of the data.[24] Therefore, the

[24]The function is increasing when the variance of damages as measured by the variance of MSK-Intensity is below 0.608 for Mathematics and below 0.328 for Spanish. This is true in over 99 percent of

variance of earthquake damages is a shifter to the variance of student types, and the main findings on the treatment effect of damage dispersion can be interpreted as the effects of the variance of peer types.

Table 4: Regression of variance of predicted types on variance of damages

| | Mathematics | | | Spanish |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Variance in earthquake damages | 0.017 | 0.042** | -0.004 | 0.040** |
| | (0.012) | (0.021) | (0.012) | (0.020) |
| Squared variance in earthquake damages | | -0.035 | | -0.061*** |
| | | (0.024) | | (0.023) |
| Observations | 2005 | 2005 | 1938 | 1938 |
| $R^2$ | 0.976 | 0.976 | 0.982 | 0.982 |

Standard errors in parentheses. Sample of classrooms in the post-earthquake cohort, in Municipal schools. Controls: classroom mean, variance, skewness and kurtosis of parental years of education (average between mother and father) and of baseline test scores; class size; number of students with family incomes above 150k CLP a month; number of female students; average of earthquake damages among classmates; damages in school town.

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

# 7 Estimation Results: Treatment Effects

## 7.1 Nonparametric Estimates of the Treatment Effects Function

The estimates of the treatment effect functions $TE(c_i)$ for Spanish and Mathematics are reported in Figure 8. The results indicate that peer effects are heterogeneous. The effect of increasing damage dispersion and, therefore, of increasing peer type variance on student test scores is heterogeneous depending on a student's type $c_i$. It is worth noting that some students benefit from an increase in peer type dispersion, indicating that results from linear models, like the ones generating Figure 5, mask considerable heterogeneity. Going from low to high $c$, the function $\widehat{TE(c)}$ is negative and then positive for Spanish test scores, while it is positive, then negative and then positive for Mathematics test scores. This means that increasing the variance of peer types has a negative impact on the test scores of middle-$c$ students, and a positive impact on the test scores of high-$c$ students, while it has a negative impact on low-$c$ Spanish students, and a positive impact on low-$c$

classrooms.

Mathematics students. Section 10.3 provides a conceptual framework to interpret these results through the lens of a game theoretical model of effort choices in the classroom. For now, it is worth noticing that peer effects are nonlinear (that is, the variance of student types matters) and heterogeneous across students.

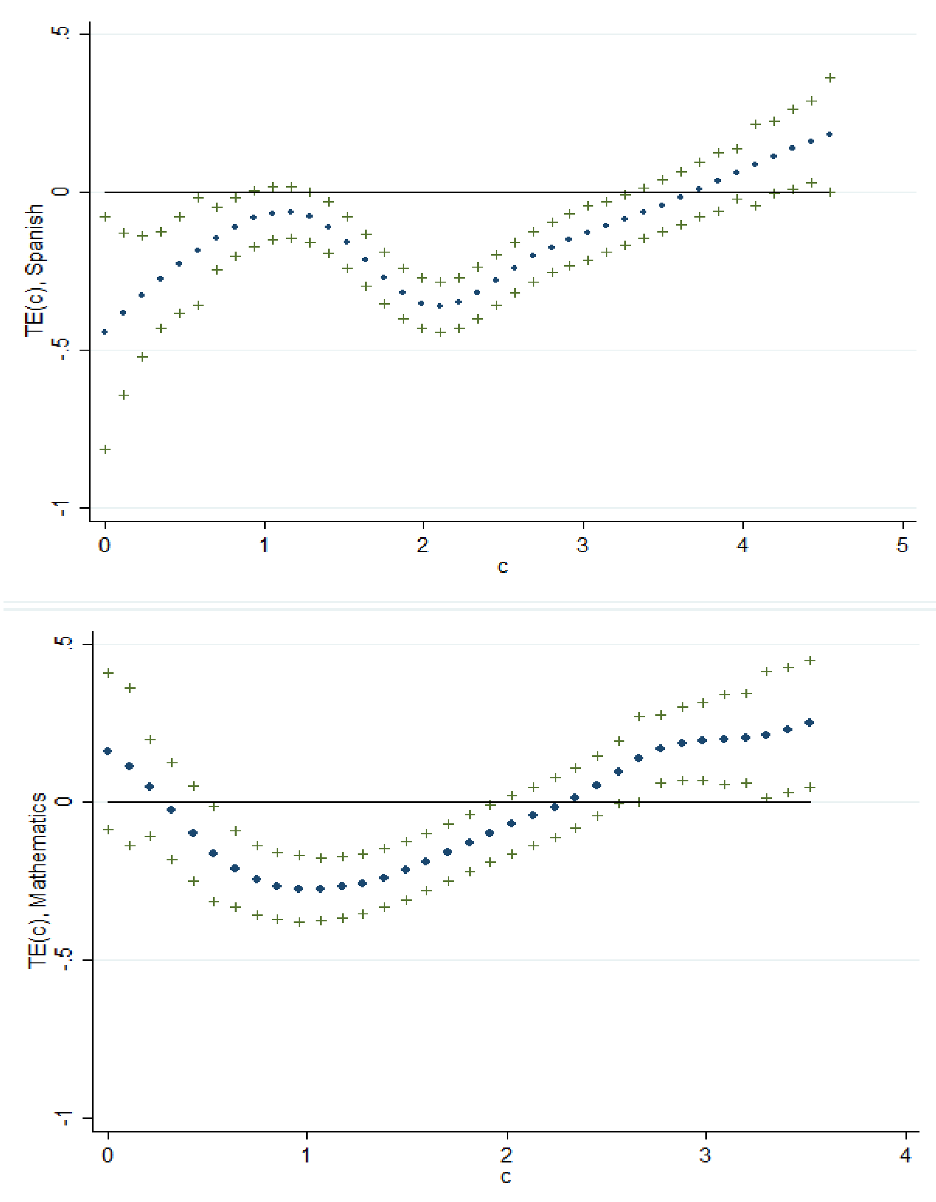


Figure 8: Estimated $\widehat{TE(c_i)}$ for Spanish (top) and Mathematics (bottom) test scores. Bounds for one-sided significance tests at the 10 percent significance level are reported. When the lower bound is above 0, we accept the hypothesis that $TE(c) > 0$, when the upper bound is below 0, we accept $TE(c) < 0$.

# 8 Model Fit

The model fit is very good, as can be seen in Table 5. This is not surprising, given the use of nonparametric techniques.

# 9 Relationship to Existing Nonlinear Difference in Differences Models

Like the semiparametric model presented here, existing nonlinear difference-in-differences models estimate flexibly the heterogeneity of treatment effects. The closest models to the one presented here are the quantile Difference-in-Differences (QDID) and Changes-in-Changes (CIC) models in Athey and Imbens (2006), and the nonlinear difference in differences models in Blundell and Costa Dias (2000), Bell, Blundell, and Van Reenen (1999), Blundell, Costa Dias, Meghir, and Van Reenen (2004), Smith and Todd (2005), Heckman, Ichimura, Smith, and Todd (1998), Abadie (2005) and D'Haultfoeuille, Hoderlein, and Sasaki (2015). However, three features distinguish the model in this paper from existing methods: nonseparability in the vector of student characteristics ($X_i$), continuous treatment, and collapsing of the covariates in $X_i$ into a single scalar $c_i$. Existing methods do not allow for all three, as explained in detail in Appendix D.1.

# 10 Mechanisms

## 10.1 Unobserved School Inputs

An earthquake is a complex disruption affecting not only students and households, but also teachers and schools. Moreover, not all teacher and school characteristics are observed. Therefore, the findings on the impact of damage dispersion may be caused by a response of unobserved school and/or classroom inputs to damage dispersion. To test whether this is the case, ideally, one would like to estimate classroom fixed effects before and after the earthquake, and explore whether the estimated fixed effects reacted systematically to earthquake damage dispersion, the treatment variable. However, this is not possible, because the econometric model already estimates (nonlinear) classroom fixed effect functions, and interprets their reaction to the earthquake damage dispersion as the treatment effect. Instead, it is possible to do the analysis at the school, rather than classroom, level. I exploit the existence of schools with multiple classrooms to estimate

Table 5: Model Fit, Test Scores

| | Mathematics | | Spanish | |
|---|---|---|---|---|
| | Actual | Model | Actual | Model |
| Pre-Earthquake Cohort | | | | |
| Overall | -.185 | -.189 | -.121 | -.123 |
| Female | -.304 | -.283 | -.050 | -.063 |
| Male | -.058 | -.089 | -.196 | -.186 |
| Female | | | | |
| Urban | -.300 | -.279 | -.052 | -.064 |
| Rural | -.322 | -.302 | -.043 | -.056 |
| Male | | | | |
| Urban | -.035 | -.066 | -.180 | -.172 |
| Rural | -.159 | -.188 | -.262 | -.249 |
| Female | | | | |
| Lower Income | -.414 | -.387 | -.148 | -.155 |
| Higher Income | -.130 | -.120 | .104 | .083 |
| Male | | | | |
| Lower Income | -.222 | -.246 | -.348 | -.328 |
| Higher Income | .155 | .116 | .001 | -.003 |
| Post-Earthquake Cohort | | | | |
| Overall | -.222 | -.228 | -.153 | -.156 |
| Female | -.307 | -.292 | -.058 | -.078 |
| Male | -.132 | -.159 | -.254 | -.239 |
| Female | | | | |
| Urban | -.302 | -.287 | -.071 | -.086 |
| Rural | -.329 | -.315 | .001 | -.039 |
| Male | | | | |
| Urban | -.120 | -.148 | -.257 | -.246 |
| Rural | -.180 | -.205 | -.242 | -.209 |
| Female | | | | |
| Lower Income | -.414 | -.388 | -.146 | -.160 |
| Higher Income | -.151 | -.151 | .071 | .042 |
| Male | | | | |
| Lower Income | -.237 | -.262 | -.351 | -.322 |
| Higher Income | -.0004 | -.0304 | -.133 | -.136 |

school fixed effects, and explore whether unobserved school inputs reacted, specifically, to the variance in earthquake damages in the classrooms.[25]

Intuitively, I estimate school fixed effects and test whether earthquake damage dispersion had any causal impact on school fixed effects, and find no evidence of this. In Appendix D.2, I derive formally the condition that, if satisfied, guarantees that the empirical findings are not driven by unobserved school inputs, and I formally introduce an empirical test for this condition, as well as the test results. In the data, I find evidence for this sufficient condition.

## 10.2 Teachers

I have shown that unobserved inputs at the *school level* are not a good candidate to explain the observed empirical findings. However, a reaction of unobserved *classroom level* inputs to damage dispersion in the classroom could be contributing to the estimated effects. The most relevant unobserved classroom input is teacher's effort. To explore this channel, I use a measure of teacher effort in the classroom that is available both before and after the earthquake: the fraction of the national curriculum covered in class during the year by the Spanish and Mathematics teachers in the sample. There are three *caveats*. First, this variable is subject to considerable non-response (35 percent for Mathematics and 30 percent for Spanish teachers) and this non-response is non-random (for example, the mean of Mathematics test scores when the variable is non-missing is 0.035 and it is $-0.066$ when it is missing). Second, it is self-reported, and there may be legitimate concerns of mis-reporting. Third, the survey question in the questionnaire for Mathematics teachers changed slightly in between cohorts. In spite of these *caveats*, it would cause concern if this (imperfect) effort measure did not pass the empirical tests presented here.
Specifically, I use this variable to perform a test similar to the one used to rule out unobserved school inputs as a driver of the empirical findings (see Proposition D.2). The main difference is that teacher effort is a *classroom level* input. Intuitively, for teacher effort to explain any part of the damage dispersion effect, it must have reacted to, specifically, damage dispersion. Any other reaction of teacher effort to the earthquake is already controlled for by term $\lambda^E$ in equation (4) and, therefore, cannot explain the estimated effects.

Formally, I use a regression like the one used to test for a reaction of school fixed effects (equation (19) in the Appendix, substituting teacher effort for the outcome variable).

[25]Any other type of reaction of unobserved school inputs to the earthquake is already accounted for by term $\lambda^E$ in equation (4), and cannot explain the empirical finings.

Regression results are reported in Table 8 in Appendix B. For both Mathematics and Spanish, I cannot reject at any conventional significance level the null that teacher effort did not react to damage dispersion, indicating that the estimated treatment effects $\hat{TE}(c)$ cannot be explained by a change in this observed measure of teacher effort.

There may be unobserved components of teacher effort not captured by the measure used here. Therefore, it is not possible to rule out teacher effort as one of the drivers of the empirical findings. However, the results presented here are consistent with teacher effort not entirely explaining the damage dispersion effects and their patterns across students.

### 10.3 Peer-to-peer Interactions: Effort Game in the Classroom

In this section I propose a conceptual framework that can help understand the empirical findings. I follow the approach adopted in Blume, Brock, Durlauf, and Jayaraman (2015) of micro-founding observed peer effects using a behavioural model.[26] I refer to a technical online document for a formal presentation of the theoretical model and derivations behind the conceptual framework, and to Mierendorff and Tincani (2018) for a closely related theoretical model tested on experimental data.[27] Here, I present the main theoretical results with the aid of graphs and intuition. The theoretical results derive directly from the results in Hopkins and Kornienko (2004). Specifically, the model used here is a variation of the model of conspicuous consumption in Hopkins and Kornienko (2004), where the assumptions on the primitives of the problem are such that all the derivations in that paper hold.

[26] Blume, Brock, Durlauf, and Jayaraman (2015) assume that student achievement is a choice. I relax this assumption by allowing for the choice to not be achievement directly, but effort, and by allowing effort to have a possibly nonlinear effect on achievement. Because effort is not observed in the data but achievement is, I assume, like in Fruehwirth (2013), that achievement is monotone increasing in effort. This allows me to derive model implications on the observed outcome variable (achievement) while having a model of effort choices. There exists empirical evidence that effort increases achievement (De Fraja, Oliveira, and Zanchi 2010, Stinebrickner and Stinebrickner 2008, Stinebrickner and Stinebrickner 2004), so this is not an unrealistic assumption. However, the theoretical predictions would hold true even in the case in which students were assumed to directly choose achievement, like in Blume, Brock, Durlauf, and Jayaraman (2015). If effort were observed, it could be used to estimate an effort game in the classrooms as well as the mapping from effort to achievement. For example, see Conley, Mehta, Stinebrickner, and Stinebrickner (2015) for a model of social interactions that uses observed study time as a measure of effort. Typically, however, effort is unobserved in large administrative datasets. Given the costs of collecting time diaries, the sample sizes of datasets that include good quality effort measures is small and cannot be combined with nonparametric techniques. For example, the sample size in Conley, Mehta, Stinebrickner, and Stinebrickner (2015) is 331 students, three orders of magnitude smaller than the sample size in this paper.

[27] See supplementary material at `http://www.homepages.ucl.ac.uk/~uctpmt1/Tincani_heterogeneous_online_supplementary_material.pdf`

Students differ in terms of ability, with higher ability students having lower cost of study effort. Cost of effort is the counterpart of student type in the econometric model, therefore, this mechanism requires this interpretation of student types. Students choose how much costly effort $e$ to exert, and effort increases achievement $y$. Student's utility is increasing in own achievement. This is a standard assumption in the literature, see, for example, Blume, Brock, Durlauf, and Jayaraman (2015), Fruehwirth (2012), and De Giorgi and Pellizzari (2013).

I introduce a novel primitive of the student's problem: I assume that students' utility is increasing in achievement rank in their reference group (e.g., classroom). This assumption says that students do not only put in effort because they derive a direct utility from achievement in absolute terms, but also because they care about obtaining higher achievement (e.g. grades) than their peers. Although there exists evidence that students display rank concerns in various settings (Tran and Zeckhauser 2012, Azmat and Iriberri 2010), the possibility that rank enters the utility function has received little attention until now. In a related paper, we show how this assumption can help us understand some pervasive but unexplained experimental results on ability tracking (Mierendorff and Tincani 2018).

This assumption generates peer effects even when spillovers are not explicitly embedded in the achievement production function. This is because when students have rank concerns, own effort depends on the distribution of peer ability and this, observationally, looks like an ability peer effect. Intuitively, a low ability student might "give up" in a classroom where all other classmates are highly skilled academically. On the other hand, she might instead engage in a healthy competition with her peers when they are of similar ability to herself.

Finally, technological spillovers are allowed to (but not necessarily assumed to) enter the achievement production function. Specifically, average peer ability is allowed to directly affect own achievement. This is a standard assumption in the peer effect literature, both theoretical and empirical. For example, the working-horse model of ability peer effects until recently has been the linear in means model (see Epple and Romano (2011) and Sacerdote (2014) for reviews of this very large literature). Because mean student type is controlled for in the econometric model, this type of technological peer effect cannot be a driver of the empirical findings. Therefore, both the theory and the empirical analysis can be agnostic about the existence and shape of technological peer effects working through mean peer ability. This adds to the generalizability of the results.

There are two theoretical results. The first one states that there is a unique symmetric Nash equilibrium of the effort game, and that, in equilibrium, achievement is strictly increasing (decreasing) in student ability (type/cost of effort). Therefore, the standard result of monotonicity of achievement in student type holds also in this novel conceptual framework. I have tested for monotonicity empirically in section 6.2 and failed to reject it. Therefore, the first model implication holds in the data.

The second theoretical result is a comparative statics result that states that when the dispersion in student types in the reference group increases (in the unimodal likelihood ratio sense, like, for example, through a mean-preserving variance increase), medium-type students perform more poorly and high-type students perform better, while low-type students perform better or worse, depending on the relative strength of the preference for achievement rank in the utility function. To make these patterns easy to understand, I present them graphically in Figure 9.

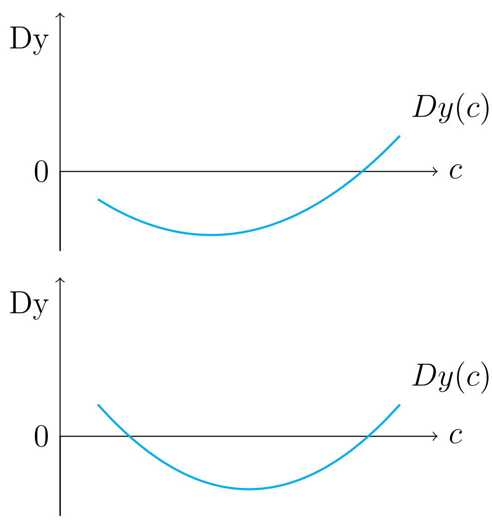


Figure 9: The function $Dy(c)$ traces the effect on achievement of increasing the dispersion (e.g.,the variance) of student type $c$, as a function of $c$. Like student types in the econometrc model, higher $c$ have higher cost of effort/lower ability. The function $Dy(c)$ can cross the x-axis once or twice. If it crosses it once (upper panel), the sequence of its signs, from low $c$ to large $c$, is $-$, $+$. If it crosses it twice (lower panel), the sequence of its signs, from low $c$ to large $c$, is $+$, $-$, $+$. As long as these crossing properties and signs are satisfied, the function $Dy(c)$ can admit any shape.

There are two intuitions for this result. First, in a world where students have rank concerns, how many students are of similar ability to oneself (technically, the density of types computed at own type) determines how easy or hard it is to improve one's own rank. The more peers there are with an ability similar to one's own, the more students can be surpassed for a unitary increase in effort. Therefore, the higher the type density at one own's type, the higher the marginal utility of effort, the higher the incentive to put in effort.

The second intuition is that a change in type variance affects the density of the type distribution differently at different points in its support. Specifically, an increase in type variance generates fatter tails: a higher density of high and low types, and a lower density of medium types.

Taken together, these two intuitions imply that high and low type students face an incentive to increase their effort when type variance increases, because there are more students at their own type level, while middle type students face an incentive to lower their effort, because there are fewer students at their type level. In fact, the theoretical model predicts, unambiguously, that low type students increase their effort and middle type students decrease it, as implied by these incentives. However, for low type students there are two simultaneous incentive: the aforementioned incentive to increase their effort in order to improve their rank, and an incentive to lower their effort to save on effort cost while not worsening their own rank, due to the fact that they now face lower competition from below (i.e., from the middle type students, who are decreasing their effort). Which effect prevails depends on how strong the preference for rank is, relative to the utility from achievement in absolute terms nett of effort cost. The model does not impose a value for this relative strength. As a result, we can let the data inform us on which effect prevails.

Comparing Figure 9 with the empirical findings reported in Figure 8, it is easy to see that the treatment effect patterns follow those predicted by the model. In particular, the patterns for Spanish resemble the top panel of Figure 9 while the patterns for Mathematics resemble the bottom panel. This is compatible with stronger rank concerns in Mathematics than in Spanish. Therefore, a model of rank concerns can explain in an intuitive and simple way the nonlinear patterns that the econometric model uncovered. This is useful, because these patterns of heterogeneity are hard to rationalize with existing models of peer effects.

# 11 Conclusions

In this paper, I combine semi-parametric techniques with a large administrative dataset and with data variation from a natural disaster to flexibly estimate peer effects in the classroom. The econometric model estimates peer effects as non-parametric functions of student characteristics. It imposes fewer restriction on the shape of peer effects than any existing method.

Estimates show that peer effects that work through the variance of peer characteristics are heterogeneous across students and follow patterns that are hard to rationalize with the existing knowledge on peer effects. I show that an effort game in the classroom, in which achievement rank enters students' payoffs, is able to rationalize the observed patterns. This demonstrates that estimating peer effects flexibly is valuable, because it can inform theories on the mechanisms behind them.

# References


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

Ahn, H., and J. L. Powell (1993): "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," *Journal of Econometrics*, 58(1), 3–29.

Altindag, A., S. Ozen, et al. (2005): "One-year follow-up study of posttraumatic stress disorder among earthquake survivors in Turkey," *Comprehensive psychiatry*, 46(5), 328–333.

Angrist, J. D., and K. Lang (2004): "Does school integration generate peer effects? Evidence from Boston's Metco Program," *American Economic Review*, 94(5), 1613–1634.

Arnott, R., and J. Rowse (1987): "Peer group effects and educational attainment," *Journal of Public Economics*, 32(3), 287–305.

Astroza, M., S. Ruiz, and R. Astroza (2012): "Damage Assessment and Seismic Intensity Analysis of the 2010 (Mw 8.8) Maule Earthquake," *Earthquake Spectra*, 28(S1), S145–S164.

Athey, S., and G. W. Imbens (2006): "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, 74(2), 431–497.

Azmat, G., and N. Iriberri (2010): "The importance of relative performance feedback information: Evidence from a natural experiment using high school students," *Journal of Public Economics*, 94(7), 435–452.

Bell, B., R. Blundell, and J. Van Reenen (1999): "Getting the unemployed back to work: the role of targeted wage subsidies," *International tax and public finance*, 6(3), 339–360.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015): "Linear social interactions models," *Journal of Political Economy*, 123(2), 444–496.

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

Blundell, R., M. Costa Dias, C. Meghir, and J. Van Reenen (2004): "Evaluating the employment impact of a mandatory job search program," *Journal of the European economic association*, 2(4), 569–606.

Booij, A., E. Leuven, and H. Oosterbeek (2016): "Ability Peer Effects in University: Evidence from a Randomized Experiment," *Review of Economic Studies, forthcoming.*

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of econometrics*, 150(1), 41–55.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013): "From natural variation to optimal policy? The importance of endogenous peer group formation," *Econometrica*, 81(3), 855–882.

Chetverikov, D. (2018): "Testing regression monotonicity in econometric models," *Econometric Theory, forthcoming.*

Cipollone, P., and A. Rosolia (2007): "Social interactions in high school: Lessons from an earthquake," *The American economic review*, pp. 948–965.

Comerio, M. C. (2013): *Housing recovery in Chile: A qualitative mid-program review.* Pacific Earthquake Engineering Research Center Headquarters at the University of California.

Conley, T., N. Mehta, R. Stinebrickner, and T. Stinebrickner (2015): "Social interactions, mechanisms, and equilibrium: Evidence from a model of study time and academic achievement," Discussion paper, National Bureau of Economic Research.

De Fraja, G., T. Oliveira, and L. Zanchi (2010): "Must try harder: Evaluating the role of effort in educational attainment," *The Review of Economics and Statistics*, 92(3), 577–597.

De Giorgi, G., and M. Pellizzari (2013): "Understanding social interactions: Evidence from the classroom," *The Economic Journal*.

Dell'Osso, L., C. Carmassi, G. Massimetti, E. Daneluzzo, S. Di Tommaso, and A. Rossi (2011): "Full and partial PTSD among young adult survivors 10 months after the L'Aquila 2009 earthquake: Gender differences," *Journal of affective disorders*, 131(1), 79–83.

D'Haultfoeuille, X., S. Hoderlein, and Y. Sasaki (2015): "Nonlinear Difference-in-Differences in Repeated Cross Sections with Continuous Treatments," .

Ding, W., and S. F. Lehrer (2007): "Do peers affect student achievement in China's secondary schools?," *The Review of Economics and Statistics*, 89(2), 300–312.

Duflo, E., P. Dupas, and M. Kremer (2011): "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *The American Economic Review*, 101, 1739–1774.

Epple, D., and R. Romano (2011): "Peer effects in education: A survey of the theory and evidence," *Handbook of social economics*, 1(11), 1053–1163.

Fruehwirth, J. C. (2012): "Can achievement peer effect estimates inform policy? A view from inside the black box," *Review of Economics and Statistics*, (00).

——— (2013): "Identifying peer achievement spillovers: Implications for desegregation and the achievement gap," *Quantitative Economics*, 4(1), 85–124.

Garlick, R. (2016): "Academic Peer Effects with Different Group Assignment Rules: Residential Tracking versus Random Assignment," Discussion paper, Working paper, World Bank Development Research Group, Washington, DC.

Ghosal, S., A. Sen, and A. W. Van Der Vaart (2000): "Testing monotonicity of regression," *Annals of statistics*, pp. 1054–1082.

Giannopoulou, I., M. Strouthos, P. Smith, A. Dikaiakou, V. Galanopoulou, and W. Yule (2006): "Post-traumatic stress reactions of children and adolescents exposed to the Athens 1999 earthquake," *European Psychiatry*, 21(3), 160–166.

GROOME, D., AND A. SOURETI (2004): “Post-traumatic stress disorder and anxiety symptoms in children exposed to the 1999 Greek earthquake,” *British Journal of Psychology*, 95(3), 387–397.

HECKMAN, J., H. ICHIMURA, J. SMITH, AND P. TODD (1998): “Characterizing Selection Bias Using Experimental Data,” *Econometrica*, 66(5), 1017–1098.

HOPKINS, E., AND T. KORNIENKO (2004): “Running to keep in the same place: consumer choice as a game of status,” *American Economic Review*, pp. 1085–1107.

HOROWITZ, J. L. (2010): *Semiparametric and nonparametric methods in econometrics*, vol. 692. Springer.

HOXBY, C. (2000): “Peer effects in the classroom: Learning from gender and race variation,” Discussion paper, National Bureau of Economic Research.

HOXBY, C. M., AND G. WEINGARTH (2005): “Taking race out of the equation: School reassignment and the structure of peer effects,” Discussion paper, Working paper.

ICHIMURA, H. (1993): “Semiparametric least squares (SLS) and weighted SLS estimation of single-index models,” *Journal of Econometrics*, 58(1), 71–120.

IMBERMAN, S. A., A. D. KUGLER, AND B. I. SACERDOTE (2012): “Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees,” *The American Economic Review*, 102(5), 2048–2082.

KANG, C. (2007): “Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea,” *Journal of Urban Economics*, 61(3), 458–495.

KREMER, M., AND D. LEVY (2008): “Peer effects and alcohol use among college students,” *The Journal of Economic Perspectives*, 22(3), 189–189.

LAVY, V., M. D. PASERMAN, AND A. SCHLOSSER (2012): “Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom,” *The Economic Journal*, 122(559), 208–237.

LUI, S., X. HUANG, L. CHEN, H. TANG, T. ZHANG, X. LI, D. LI, W. KUANG, R. C. CHAN, A. MECHELLI, ET AL. (2009): “High-field MRI reveals an acute impact on brain function in survivors of the magnitude 8.0 earthquake in China,” *Proceedings of the National Academy of Sciences*, 106(36), 15412–15417.

Lyle, D. S. (2009): "The effects of peer group heterogeneity on the production of human capital at West Point," *American Economic Journal: Applied Economics*, pp. 69–84.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 60(3), 531–542.

Mierendorff, K., and M. M. Tincani (2018): "Explaining Experimental Evidence on Ability Peer Effects with a Theory of Rank Concerns," *mimeo, UCL*.

Pagan, A., and A. Ullah (1999): *Nonparametric econometrics*. Cambridge university press.

Sacerdote, B. (2008): "When the saints come marching in: Effects of Hurricanes Katrina and Rita on student evacuees," Discussion paper, National Bureau of Economic Research.

——— (2014): "Experimental and quasi-experimental analysis of peer effects: two steps forward?," *Annual Review of Economics*, 6(1), 253–272.

Sacerdote, B. I. (2001): "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 116(2), 681–704.

Smith, J. A., and P. E. Todd (2005): "Does matching overcome LaLonde's critique of nonexperimental estimators?," *Journal of econometrics*, 125(1), 305–353.

Stinebrickner, R., and T. R. Stinebrickner (2004): "Time-use and college outcomes," *Journal of Econometrics*, 121(1-2), 243–269.

——— (2006): "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds," *Journal of public Economics*, 90(8), 1435–1454.

——— (2008): "The causal effect of studying on academic performance," *The BE Journal of Economic Analysis & Policy*, 8(1).

Tran, A., and R. Zeckhauser (2012): "Rank as an inherent incentive: Evidence from a field experiment," *Journal of Public Economics*, 96(9), 645–650.

Vigdor, J., and T. Nechyba (2007): "Peer effects in North Carolina public schools," *Schools and the equal opportunity problem*, pp. 73–101.

Whitmore, D. (2005): "Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment," *American Economic Review, P&P*, pp. 199–203.

Zimmerman, D. J. (2003): "Peer effects in academic outcomes: Evidence from a natural experiment," *Review of Economics and Statistics*, 85(1), 9–23.

# A Technical Appendix: Estimation

## A.1 Algorithm for the Estimation of the Semi-parametric Single-Index Model

1. Normalize to a constant one of the elements of $\theta$, because only the ratios among the components of $\theta$ are identified. I normalize to -1 the coefficient on lagged test score ($\theta_1$).

2. Make an initial guess for all the other elements of $\theta$.

3. Form $c_i$ $\forall i$ according to $c_i = -y_{i,t-1} + \theta_2 peduc_i + \theta_3 income_i + \theta_4 female_i$ if $i$ belongs to the pre-earthquake cohort, and $c_i = -y_{i,t-1} + \theta_2 peduc_i + \theta_3 income_i + \theta_4 female_i + \theta_5 I_i + \theta_6 I_i X_i^-$ if $i$ belongs to the post-earthquake cohort. Only household income and student gender are interacted with $I_i$.

4. Estimate $E(y_i|c, r; \theta)$ $\forall r$ by Nadaraya-Watson kernel regression with weights $w_i$:

$$\hat{h}_r(c;\theta) = \frac{\sum_{i\in r} w_i K\left(\frac{c_i - c}{b}\right) y_i}{\sum_{i\in r} w_i K\left(\frac{c_i - c}{b}\right)}$$

   with a standard normal Kernel: $K(\psi) = (2\pi)^{-\frac{1}{2}} exp(-0.5\psi^2)$ and optimal bandwidth $b = 1.06\hat{\sigma}_c n^{-1/5}$, minimizing the Approximated Mean Integrated Squared Error (AMISE).[28] The weights $w_i$ are such that only observations $i$ where the p.d.f. of $c$ at $c_i$ exceeds a small positive number are used (see Ichimura (1993) and Horowitz (2010)). Observation $i$ is excluded from the calculation of $\hat{h}$ at $c_i$.

5. Compute the sum of squared residuals in each $r$ at the current guess for $\theta$: $SSR_r(\theta) = \sum_{i\in r} w_i (y_i - \hat{h}_r(c_i;\theta))^2$. The weights are the same as those used in the kernel estimator of $h$.

6. Update guess for $\theta$ using Generating Set Search algorithm (HOPSPACK).

7. Repeat steps 1-6 until convergence to the minimizer of $\sum_r SSR_r(\theta)$.

Notice that $SSR(\theta)$ is computed in each classroom $r$, and its sum over classrooms is minimized. The parameter $\theta$ is restricted to be identical in all classrooms.

[28] The MISE is equal to $E\{\int [\hat{h}(c) - h(c)]^2 dx\} = \int \left[(Bias\hat{h})^2 + V(\hat{h})\right] dc$, and AMISE substitutes the expressions for the bias and variance of $\hat{h}$ with approximations. See Pagan and Ullah (1999), p. 24.

## A.2 Standard errors of the treatment effect function

The standard errors of $\widehat{TE(c)}$ cannot be easily bootstrapped for computational reasons.[29] Instead, I use the result in Ichimura (1993), who proves that the asymptotic variance of $\hat{h}_r(c)$ in the appropriately weighted semiparametric single-index model above is identical to the asymptotic variance of a non-parametric conditional mean estimator. The variance of such estimator is $V(\hat{h}_r(c)) = \frac{\sigma_r^2}{n_r h_r f_r(c)} \int K^2(\psi)d\psi + o(n^{-1}b_r^{-1})$, where $\sigma^2$ is the variance of $\epsilon_{ir}$, $b_r$ is the bandwidth, $n_r$ is the size of classroom $r$ (on average this is around 30), and $f_r(c)$ is the density at $c$ in classroom $r$. The kernel $K(\cdot)$ is the normal kernel, resulting in $\int K^2(\psi)d\psi = 0.2821$. I estimate the asymptotic variance of $\hat{h}_r(c)$ $\forall r$ on a fine grid for $c$. I substitute $f(c)$ with its kernel estimator, and $\sigma_r^2$ with its estimator obtained by averaging the squared residuals in each classroom: $\hat{\sigma}_r^2 = \frac{\sum_{i\in r}(y_i - \hat{y}_i)^2}{n_r - 1}$. I assume that the covariances between the $\hat{h}_r(c)$ belonging to different classrooms $r$ are zero $\forall c$, and I obtain the following expression for the variance of $\widehat{TE(c)}$:

$$V\left(\widehat{TE(c)}\right) = \sum_{r=1}^{N^{pre}-1} \sum_{r'=l+1}^{N^{pre}} \sum_{s=1}^{N^{post}-1} \sum_{s'=s+1}^{N^{post}} \kappa_{rr'ss'}^2 \left(V\left(\hat{h}_s^{post}(c)\right) + V\left(\hat{h}_{s'}^{post}(c)\right) + V\left(\hat{h}_r^{pre}(c)\right) + V\left(\hat{h}_{r'}^{pre}(c)\right)\right).$$

The weights $\kappa_{rr'ss'}$ are given by:

$$\kappa_{rr'ss'} = \frac{\omega_{rr'ss'}}{\sum_{r=1}^{N^{pre}-1} \sum_{r'=r+1}^{N^{pre}} \sum_{s=1}^{N^{post}-1} \sum_{s'=s+1}^{N^{post}} \omega_{rr'ss'}}$$

where $\omega_{rr'ss'}$ is defined in equation 10 below.

## A.3 Kernel Weighting

To ensure that the classrooms are similar, I assign increasing weights to quadruplets that are more similar in terms of $W$ and $\delta$. I construct weights using multivariate standard normal kernel functions. As in the main text, let $ss'$ index a pre-earthquake classroom pair, and $rr'$ a post-earthquake classroom pair. Letting $t = r, r', s, s'$ I assign the weight $\frac{1}{b}k\left(\frac{W_t - W_{t'}}{b}\right)$ to each of the pairs $tt' \in \{rr', ss', rs\}$. This ensures similarity between pairs within ($tt' = rr', ss'$) and across ($tt' = rs$) cohorts.[30] Finally, I build a weight that is declining in $|\delta_{ss'} - \delta_{rr'}|$, to guarantee that the pre- and post-earthquake pairs differ in terms of geographic dispersion $\delta$ in a similar way: $\frac{1}{b_\delta}k\left(\frac{\delta_{ss'} - \delta_{rr'}}{b_\delta}\right)$. The weight for the

[29] This would require submitting around 4,000 jobs of duration 72 hours each.

[30] I use a unique bandwidth $b$. Following Pagan and Ullah (1999), I normalize the elements in $W_t$ so that they all have the same standard deviation and using a unique bandwidth is admissible.

quadruple, $\omega_{rr'ss'}$, is the product of these four kernel weights:

$$\omega_{rr'ss'} = d_{rr'ss'} \frac{1}{b_\delta} k \left( \frac{\delta_{ss'} - \delta_{rr'}}{b_\delta} \right) \prod_{tt' \in \{rr', ss', sr\}} \frac{1}{b} k \left( \frac{W_t - W_{t'}}{b} \right) \tag{10}$$

where $d_{rr'ss'}$ is a dummy variable equal to one if $\delta_{rr'} > 0$ and $\delta_{ss'} > 0$, zero otherwise.

# B Additional Tables and Figures

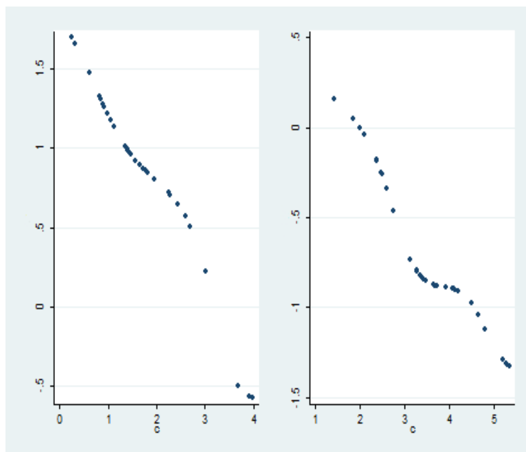

Figure 10: Plots of estimated $h$ functions in two classrooms.

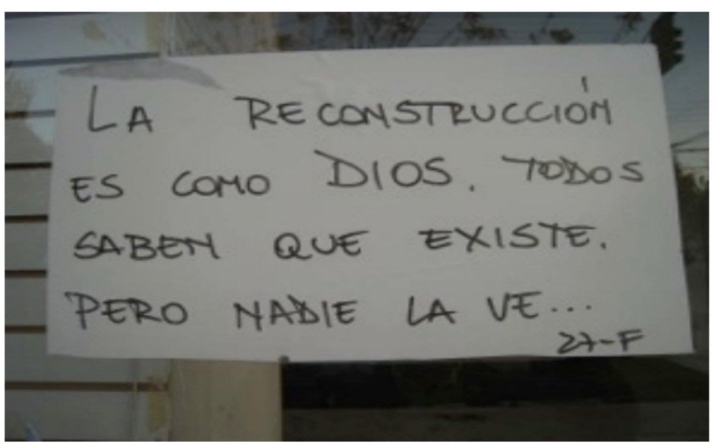


Figure 11: *Source*: Comerio (2013). Handmade sign found in Cauquenes, Chile, on February 2, 2012. Translation: "Reconstruction is like God. Everyone knows it exists, but nobody has seen it."

Table 6: Impact of variance of future earthquake intensity in the pre-earthquake cohort. (Municipal schools sample)

| | (1) Math TS | (2) Math TS | (3) Spanish TS | (4) Spanish TS |
|---|---|---|---|---|
| Variance of MSK-Intensity in the classroom | $0.232^{***}$ | $0.137^{***}$ | $0.271^{***}$ | $0.178^{***}$ |
| | (0.030) | (0.041) | (0.032) | (0.044) |
| Lagged TS | $0.694^{***}$ | $0.672^{***}$ | $0.713^{***}$ | $0.699^{***}$ |
| | (0.003) | (0.004) | (0.003) | (0.005) |
| Controls | No | Yes | No | Yes |
| Observations | 45,814 | 26,145 | 46,127 | 25,628 |

Standard errors in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Controls: mean intensity in the classroom (in all columns), whether the student lives in the same town where the school gender, mother's education, father's education, household income, intensity of earthquake in home town and in school town, class size, whether the Math or Spanish teacher is female, has a postgraduate degree, has a permanent contract, her tenure at the school, her teaching experience. A constant is always included.

Table 7: Probit regression, marginal probability estimates reported. Dependent variables: being at the top (1) or bottom (2) third of the distribution of elicited cost of effort.

| | top 33 percent (1) | bottom 33 percent (2) |
|---|---|---|
| Lagged Math TS | $-0.055^{***}$ | $0.100^{***}$ |
| | (0.002) | (0.002) |
| Seismic intensity at student's home | $0.013^{***}$ | $-0.011^{***}$ |
| | (0.002) | (0.003) |
| SES Controls | Yes | Yes |
| Observations | 46,059 | 46,059 |

Standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

SES controls: father's and mother's education, household income. Post-earthquake sample.

# C Testing Monotonicity of $h(c)$

The procedure that I use is an application of Chetverikov (2018). It would be computationally unfeasible to perform the test in all classrooms. Therefore, I create 72 categories of classrooms, containing approximately 60 classrooms each, that have similar distributions of $c$, and test monotonicity within each category. The monotonicity of $h$ is tested within each one of these categories. In all categories, the null hypothesis that the $h$ function is monotonically decreasing in $c_i$ is not rejected at the $\alpha = 0.10$ significance level.

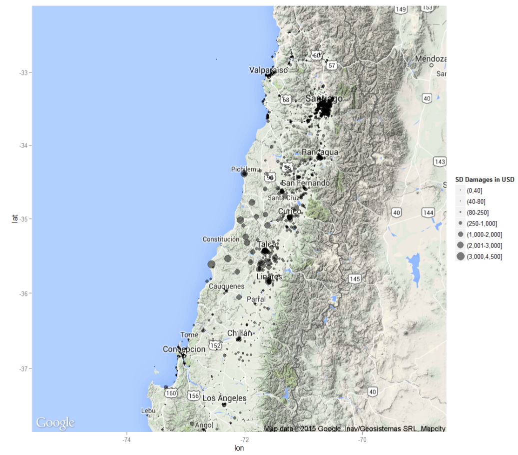


Figure 12: Schools where not all students are affected equally by the earthquake, and within school standard deviation in damages.

Consider the i.i.d. sample $\{c_i, -y_i\}_{1\leq i\leq n_l}$, where $n_l$ is the size of the $l^{th}$ classroom category.[31] Let $c_i$ and $c_j$ be a pair of observations for $c$. The test function within each category $l$ is defined as:

$$b(s) = b(\{c_i, -y_i\}, s) = \frac{1}{2}\sum_{1\leq i,j\leq n}(-y_i + y_j)sign(c_j - c_i)Q(c_i, c_j, s)$$

where I dropped the $l$ subscript for convenience, and where $Q(c_i, c_j, s)$ is a weighting function indexed by $s \in S$. To each $s$ corresponds a choice of point $c$ and bandwidth $h$ for the following specification of the weighting function:

$$Q(c_1, c_2, (c, h)) = K\left(\frac{c_1 - c}{k}\right)K\left(\frac{c_2 - c}{k}\right)$$

where $K(u) = 0.75(1 - u^2)$ if $-1 < u < 1$, and $= 0$ otherwise, and where $k = \frac{1}{2}n_l^{-\frac{1}{5}}$.[32] I let $c$ take on 100 values, which are equally spaced points going from the smallest to the largest observed value of $c_i$ in the population. As a result, there are 100 weighting functions for each classroom category $l$.

Conditional on $\{c_i\}$, the variance of $b(s)$ is given by:

$$V(s) = V(\{c_i\}, \{\sigma_i\}, s) = \sum_{1\leq i\leq n}\sigma_i^2\left(\sum_{1\leq j\leq n}sign(c_j - c_i)Q(c_i, c_j, s)\right)^2 \tag{11}$$

[31] $y_i$ is replaced by $-y_i$, and $h$ will be replaced by $-h$, because this procedure tests that $-h$ is increasing, which is equivalent to testing that $h$ is decreasing.

[32] This is the value for the bandwidth recommended in Ghosal, Sen, and Van Der Vaart (2000).

Table 8: Is measured teacher productivity in the classroom a channel behind the empirical findings?

| Dependent Variable: | Curriculum Covered | |
|---|---|---|
| | Mathematics | Spanish |
| | (1) | (2) |
| $\beta_1$ (st. dev.) | -0.00246 | 0.00525 |
| | (0.0412) | (0.0355) |
| $\beta_2$ (cohort) | -0.0105 | 0.0191 |
| | (0.0439) | (0.0357) |
| $\beta_3$ (st. dev.$\times$ treatm. dummy) | 0.0364 | 0.0264 |
| | (0.0513) | (0.0435) |
| Curriculum Covered | | |
| Constant $\beta_0$ | $0.555^{***}$ | $0.638^{***}$ |
| | (0.0430) | (0.0337) |
| Controls | Yes | Yes |
| Observations | 2,902 | 3,187 |

Standard errors in parentheses. $^{+}$ $p < 0.105$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

*Notes.* Sample of Municipal schools.

Models (1) and (2) are regression (19). The unit of observation is the classroom. Included regressors are the classroom means of: MSK-intensity, also interacted with the cohort dummy, lagged Math and Spanish test scores, mother's and father's education, income; and a set of teacher and classroom characteristics.

where $\sigma_i = (E[\epsilon_i^2|c_i])^{\frac{1}{2}}$ and $\epsilon_i = -y_i - (-h(c_i))$. Following Chetverikov (2018), I use the residual $\hat{\epsilon_i} = -y_i - (-h(c_i))$ as an estimator for $\sigma_i$, and obtain the estimated conditional variance of $b(s)$ by substituting $\sigma_i^2$ with $\hat{\sigma}_i^2$ in equation 11. The test statistic is given by:

$$T = T(\{c_i, -y_i\}, \{\hat{\sigma}_i\}, S) = \max_{s \in S} \frac{b(\{c_i, -y_i\}, s)}{\sqrt{\hat{V}(\{c_i\}, \{\hat{\sigma}_i\}, s)}}.$$

Large values of $T$ indicate that the null hypothesis that $-h$ is increasing is violated.

To simulate the critical values, I adopt the plug-in approach. The goal is to obtain a test of level $\alpha$. Let $\{\xi_i\}$ be a sequence of $B$ independent $N(0,1)$ random variables that are independent of the data. Let $-y_{i,b}^* = \hat{\sigma}_i \xi_{i,b}$ for each $b = 1, B$ and $i = 1, n$, where $\hat{\sigma}_i = \hat{\epsilon}_i$. For each $b = 1, B$, calculate the value $T_b^*$ of the test statistic using the sample $\{c_i, -y_{i,b}^*\}_{i=1}^n$. The plug-in critical value $c_{1-\alpha}$ is the $(1-\alpha)$ sample quantile of $\{T_b^*\}_{b=1}^B$.

Table 9: Values of test statistics and critical values for test of monotonicity at the $\alpha = 0.10$ significance level, by classroom category. 8 randomly selected categories.

| | Mathematics | | Spanish | |
|---|---|---|---|---|
| Classroom Category | Test statistic | Critical Value | Test statistic | Critical Value |
| Pre-earthquake classrooms | | | | |
| 1 | 2.2874460E-02 | 4.60e+19 | 4.0707965E-03 | 1.05e+19 |
| 2 | 6.2840671E-04 | 1.08e+19 | 1.6759724E-03 | 7.75e+18 |
| 3 | 3.6209350E-04 | 4.98e+18 | 1.0020613E-03 | 1.79e+19 |
| 4 | 3.9056635E-03 | 1.92e+19 | 2.2328943E-03 | 1.97e+19 |
| Post-earthquake classrooms | | | | |
| 5 | 1.0598215E-03 | 6.01e+18 | 3.8213478E-04 | 1.63e+19 |
| 6 | 2.7184933E-03 | 1.41e+19 | 1.1514544E-03 | 1.32e+19 |
| 7 | 4.1919011E-03 | 4.22e+19 | 1.3525186E-03 | 1.19e+19 |
| 8 | 1.7282768E-03 | 1.22e+19 | 3.4069275E-03 | 1.42e+19 |

In all classroom categories, the test statistic is below the critical value. Therefore, the null hypothesis that $h(c)$ is monotonically decreasing is not rejected.

# D Technical Appendix: Robustness

## D.1 Nonlinear D-i-D Models

The model in this paper accommodates the continuity of the treatment variable. It exploits it by defining treatment status within pairs of classrooms, with treatment defined as an increase, of any amount, in the treatment variable. Existing non-linear difference-in-differences models cannot be used in this context because they do not accommodate continuous treatment. For example, the changes-in-changes and quantile-difference-in-differences (QDID) models in Athey and Imbens (2006) compare outcome distributions across multiple groups and time periods, however, treatment status is a binary variable. Similarly, continuity of treatment is not accommodated within the framework of the non-linear D-i-D models based on propensity score matching in Blundell and Costa Dias (2000), Bell, Blundell, and Van Reenen (1999), Blundell, Costa Dias, Meghir, and Van Reenen (2004), Smith and Todd (2005), and Heckman, Ichimura, Smith, and Todd (1998). On the other hand, the multi-level treatment case in Abadie (2005) and the model in D'Haultfoeuille, Hoderlein, and Sasaki (2015) accommodate continuous treatment. However, they do not allow the researcher to collapse the student covariates in $X_i$ into a single scalar, $c_i$. In fact, no existing method allows for this feature.

Nonseparability in $X$ is due to the fact that student type $c_i$ is determined differently in the pre- and post-earthquake cohort. As a result, the outcome function cannot be expressed as a component that only depends on damage dispersion status (high or low) and one that only depends on cohort, conditional on $X$. This is shown in detail below. As described in the main text, this feature of the model is important to make the correct comparisons *across cohorts*, that is, between treated post-earthquake students and control pre-earthquake students. Separability is assumed in the QDID model, as well as in the models that combine matching with differences in differences, that is, Blundell and Costa Dias (2000), Bell, Blundell, and Van Reenen (1999), Blundell, Costa Dias, Meghir, and Van Reenen (2004), Smith and Todd (2005), Heckman, Ichimura, Smith, and Todd (1998), and Abadie (2005). These models, therefore, are not well-suited in this context.

To exemplify non-separability, I describe it within the framework of a hypothetical QDID model applied to the context of this paper. Such a model would assume additive separability of the outcome function. Specifically, the outcome function in the absence of treatment (which is used to build the distribution of counterfactual outcomes for treated individuals) would be assumed to be: $Y^N = h(U, G, T, X) = h^G(U, G, X) + h^T(U, T, X)$ where $G$ indicates dispersion (high or low) and $T$ indicates cohort, $U$ is an individual's unobservable, and $X$ is a vector of individual characteristics.[33] That is, function $h$ would be composed of an outcome function that only depends on dispersion status $h^G$ and one that only depends on cohort $h^T$, conditional on $X$. In the context of this paper this assumption is not satisfied if the vector $X$ enters the outcome function as an index $c_i$ and if the same vector $X$ contributes to generate a student's type $c_i$ differently in the pre- and post-earthquake cohort. This is the case when damage to a student's home has no effect on students' type before the earthquake (because damage has not occurred yet), but it does after the earthquake. Formally, the identifying assumption for QDID fails if what enters the outcome function is an index $c$ which is a cohort-specific function of $X$. The data generating process in the absence of treatment would then be $Y^N = h(U, G, T, c^T(X)) = h^G(U, G, c^T(X)) + h^T(U, T, c^T(X))$, where it is clear that the first function depends on both $G$ and $T$ and, therefore, QDID would be misspecified, because additive separability would not be satisfied. The model presented in this paper relaxes the assumption of additive separability conditional on $X$. Additive separability holds only conditional on $c$, a weaker assumption.

[33] I am switching the notation with respect to, for example, the notation in Athey and Imbens (2006). In particular, the control group here is $T = 0$, the pre-earthquake cohort, rather than $G = 0$. The time-trend in Athey and Imbens (2006) corresponds to the geographic dispersion effect here. All results are unchanged.

## D.2 Robustness to School Fixed Effects

**Proposition D.1 Robustness to school×cohort fixed effects.** *Under the condition below, the point estimator of the treatment effect* $\hat{TE}(c)$ *in equation* (8) *is an unbiased estimator of the causal effect of damage dispersion on achievement in the presence of school×cohort fixed effects, for all values of* $c_i$. *Moreover, it converges in probability to the true causal effect as the number of schools in the sample goes to infinity. The condition provides that:*

$$\begin{aligned} & \Big[E[\alpha_{RT}|T_M = 1, G_{RR'} = 1, c, W, \delta] - E[\alpha_{RT}|T_R = 1, G_{RR'} = 0, c, W, \delta]\Big] = \\ = \; & \Big[E[\alpha_{RT}|T_R = 0, G_{RR'} = 1, c, W, \delta] - E[\alpha_{RT}|T_R = 0, G_{RR'} = 0, c, W, \delta]\Big] \quad \forall \delta, c, W, \end{aligned} \tag{12}$$

*where* $\alpha_{RT}$ *is the school×cohort fixed effect,* $R$ *is a school indicator,* $T_R$ *is a cohort indicator, and the other variables are defined in the main text.*

**Proof of Proposition D.1.**

Suppose that the true Data Generating Process is the nonlinear model in (5), augmented with unobserved school×cohort effects $\alpha_{RT}$:

$$\begin{aligned} Y_{irR} &= H(c^{T_r}(X_i), W_r, T_r, \sigma_r^2, R) + \epsilon_{irR} \\ &= h(c^{T_r}(X_i), W_r, T_r, \sigma_r^2) + \alpha_{RT} + \epsilon_{irR} \\ &= \phi(c_i, W_r) + \lambda^{GD}(c_i, W_r, \sigma_L^2) + T_r \cdot \Big[\lambda^E(c_i, W_r) + \lambda^{DD}(c_i, W_r, \sigma_L^2)\Big] + \\ &\quad + G_{r,r'} \cdot \lambda^{GD}(c_i, W_r, \delta) + T_r \cdot G_{rr'} \cdot \lambda^{DD}(c_i, W_r, \delta) + \alpha_{RT} + \epsilon_{irR} \end{aligned} \tag{13}$$

To keep track of the school each classroom is in, I use an upper case letter for the school, so, for example, classroom $r$ is in school $R$, classroom $r'$ is in school $R'$, et cetera. The fixed effect $\alpha_{RT}$ is identified from schools with multiple classrooms. Consider one quadruplet of classrooms, $r, r'$ from the pre-earthquake cohort, $l, l'$ from the post-earthquake cohort, sharing the same $W$ and with $G_{rr'} = 1$, $G_{ss'} = 1$, and $\delta_{rr'} = \delta_{ss'} = \delta$. I drop the $T$ subscript from the fixed effect because the school index now also uniquely identifies the cohort. Conditional on a point $c$, on $W$ and on $\delta$, the double difference now yields:

$$\begin{aligned} \Big(H_r(c_i) - H_{r'}(c_i)\Big) - \Big(H_s(c_i) - H_{s'}(c_i)\Big) &= \Big(\lambda^{GD}(c_i; W, \delta) + \lambda^{DD}(c_i; W, \delta)\Big) - \lambda^{GD}(c_i; W, \delta) + \\ &\quad + (\alpha_R - \alpha_{R'}) - (\alpha_S - \alpha_{S'}) \\ &= \lambda^{DD}(c_i; W, \delta) + (\alpha_R - \alpha_{R'}) - (\alpha_S - \alpha_{S'}) + \zeta_{iRR'SS'} \end{aligned} \tag{14}$$

There are two cases two consider. First, when $R = R'$ and $S = S'$, that is, when within cohorts the pairs of classrooms are in the same school, school×cohort fixed effects cancel out. No restrictions on the fixed effects would have to be imposed. However, only 37 percent of schools in the sample have more than one classroom, hence, it is reasonable to expect that there is only a small number of within cohort matched classroom pairs that belong to the same school. Therefore, I consider the properties of the nonlinear model in the more empirically relevant case that $R \neq R'$ or $S \neq S'$ or both.

In this case, it is easy to see that the conditional expectation of the double difference in 14, $E\Big[\Big(H_r(c_i) - H_{r'}(c_i)\Big) - \Big(H_s(c_i) - H_{s'}(c_i)\Big)|c, W, \delta\Big]$, is equal to the conditional damage dispersion effect, $\lambda^{DD}(c_i; W, \delta)$, if and only if

$$E\Big[(\alpha_R - \alpha_{R'}) - (\alpha_S - \alpha_{S'})|c, W, \delta\Big] = 0 \quad \forall c, W, \delta, \tag{15}$$

where the expectation is taken with respect to the distribution of school×cohort fixed effects. In turn, when the condition in (15) is true, equation (7) identifies the treatment effect, $TE(c)$, and the expectation of the sample mean in (8) is equal to $TE(c)$ for all values of $c$. That is, the nonlinear estimator is unbiased. To see why, notice that when the true DGP includes fixed effects, equation (8) is equivalent to:

$$\begin{aligned} T\hat{E}(c) \quad = \quad & \frac{\sum_{s=1}^{N^{pre}-1}\sum_{s'=s+1}^{N^{pre}}\sum_{r=1}^{N^{post}-1}\sum_{r'=r+1}^{N^{post}}\omega_{rr'ss'}(\Delta^{post}\hat{h}_{rr'}(c;W,\delta) - \Delta^{pre}\hat{h}_{ss'}(c;W,\delta))}{\sum_{s=1}^{N^{pre}-1}\sum_{s'=s+1}^{N^{pre}}\sum_{r=1}^{N^{post}-1}\sum_{r'=r+1}^{N^{post}}\omega_{rr'ss'}} + \\ & + \frac{\sum_{s=1}^{N^{pre}-1}\sum_{s'=s+1}^{N^{pre}}\sum_{r=1}^{N^{post}-1}\sum_{r'=r+1}^{N^{post}}\omega_{rr'ss'}(\hat{\alpha}_R - \hat{\alpha}_{R'}) - (\hat{\alpha}_S - \hat{\alpha}_{S'})}{\sum_{s=1}^{N^{pre}-1}\sum_{s'=s+1}^{N^{pre}}\sum_{r=1}^{N^{post}-1}\sum_{r'=r+1}^{N^{post}}\omega_{rr'ss'}}. \end{aligned}$$

The second line is the empirical counterpart of $E\Big[E\Big[(\alpha_R - \alpha_{R'}) - (\alpha_S - \alpha_{S'})|c, W, \delta\Big]\Big]$, and, under condition (15), its expectation is equal to 0 by the central limit theorem. Additionally, as the number of schools goes to infinity, the second line of the expression above converges to zero, i.e., to its population counterpart under (15), by the weak law of large numbers for independent and not identically distributed random variables.[34] Therefore, when there are school×cohort fixed effects in the DGP, condition (15) is a necessary and sufficient condition for the identification of the conditional treatment effect

[34] The elements of the average in the second line of the expression above are not identically distributed because the sampling variance of each element depends on the school size, that is, on the sample size on which the fixed effects that enter the double difference are calculated. It is the number of schools that must go to infinity for convergence because the expectation in (15) is taken with respect to the school fixed effect distribution.

$\lambda^{DD}(c_i; W, \delta)$, and it is sufficient for the identification of the unconditional treatment effect, $TE(c)$. Finally, using the definitions of $R, R', S$ and $S'$, condition (15) can be rewritten as it appears in the main text:

$$\begin{aligned} &\Big[E[\alpha_{RT}|T_R = 1, G_{RR'} = 1, c, W, \delta] - E[\alpha_{RT}|T_R = 1, G_{RR'} = 0, c, W, \delta]\Big] = \\ = \quad &\Big[E[\alpha_{RT}|T_m = 0, G_{RR'} = 1, c, W, \delta] - E[\alpha_{RT}|T_R = 0, G_{RR'} = 0, c, W, \delta]\Big] \quad \forall c, W, \delta. \end{aligned}$$

In addition, the school effects may have heterogeneous impacts on students. To see why, replace $\alpha_{RT}$ with a function $\alpha_{RT}(c)$ and all derivations above hold true.

**Proposition D.2 Identification test in the presence of fixed effects.** *If the conditional expectation of the school×cohort fixed effect is a linear function of $\sigma_r^2$ and $T_R$, with $W$ and $c$ entering in an additively separable way, then a sufficient condition for identification of the treatment effect $TE(c)$ is that $\beta_3 = 0$ in the following equation:*

$$\alpha_{RT} = \beta_0 + \beta_1\sigma_r^2 + \beta_2 T_R + \beta_3\sigma_r^2 T_R + g(W_r, c_i; \beta_4) + \epsilon_{rR} \quad \forall r \in R. \tag{16}$$

**Proof of Proposition D.2.**

The sufficient condition for identification with fixed effects in (12) must be true for every $\delta$. In particular, it must be true for $\delta \to 0$, in which case, if the conditional expectation of the fixed effect is differentiable in $\sigma_r^2$, it must be that

$$\frac{\partial E[\alpha_R|T_R = 1, c, W, \sigma_r^2]}{\partial\sigma_r^2} = \frac{\partial E[\alpha_R|T_R = 0, c, W, \sigma_r^2]}{\partial\sigma_r^2} \quad \forall r \in R, \forall\sigma_r^2, c, W. \tag{17}$$

A specification for the conditional expectation of the fixed effects that is useful for testing is the special case in which this expectation is linear in $\sigma^2$ and $T$. In this case, condition (17) is also sufficient for identification of $TE(c)$ when there are school×cohort fixed effects in the DGP.[35] A further simplifying (but not necessary) assumption is that $W$ and $c$ enter in an additively separable way. Condition (17) under this model of fixed effects is equivalent to $\beta_3 = 0$ in:

$$\alpha_{RT} = \beta_0 + \beta_1\sigma_r^2 + \beta_2 T_R + \beta_3\sigma_r^2 T_R + g(W, c; \beta_4) + \epsilon_{rR}. \tag{18}$$

[35] To see why, notice that if $E[\alpha_R|T_R, c, W, \sigma_r^2]$ is linear in $\sigma_r^2$ and it has the same slope under $T_R = 1$ and $T_R = 0$, then condition (12) is satisfied for all values of $\delta$ and not only for $\delta \to 0$.

Therefore, $\beta_3 = 0$ is a sufficient condition for identification in the presence of fixed effects, when the fixed effects follow the specification in (18).

To verify if this condition is satisfied in the data, I estimate damage dispersion effects in a linear D-i-D model (for numerical tractability reasons) with the addition of school fixed effects, and compute predicted fixed effects $\hat{\alpha}_{RT}$ using the estimated parameters.[36] I then verify if the condition for identification under linearity of the fixed effects in (16) is rejected in the data by estimating the following linear model:

$$\hat{\alpha}_{RT} = \beta_0 + \beta_1\sigma_r^2 + \beta_2 T_R + \beta_3\sigma_r^2 T_R + \beta_4 W_R + \beta_5\tilde{c}_R + \epsilon_R \tag{19}$$

where, for simplicity, I have replaced $g(W, c; \beta_4)$ with a linear function, and where $\tilde{c}_R$ are student characteristics aggregated at the school level like, for example, average income and average parental education.[37] A t-test on the significance of the $\hat{\beta}_3$ estimated coefficient cannot reject $\beta_3 = 0$, for both Mathematics and Spanish (p-values: 0.171 and 0.682 respectively). Estimation results are reported in Table 10. Therefore, the identifying assumption under fixed effects is not rejected.

[36] However, in the linear model only the average fixed effect across students in the same school can be estimated. That is, the heterogeneity of the school effect across students cannot be captured in the linear model. Therefore, I am testing robustness of $\hat{TE}(c)$ to a standard kind of school fixed effects (that is, a constant school effect for all students in the same school). To keep a close correspondence between the semiparametric and the linear models, the linear regression used to compute the fixed effects contains the same set of $W$ and $X$ characteristics as the non-linear model.

[37] For each school there are as many equations as there are classrooms, because $\sigma_r^2$ is classroom specific. Therefore, the most appropriate model is a seemingly unrelated regression (SUR). However, given the high correlation in $\sigma_r^2$ across classrooms within the same school, for simplicity, I compute overall school level $\sigma_R^2$, and estimate the regression in 19 using $\sigma_R^2$ in place of $\sigma_r^2$. Because the SUR estimator is expected to have a higher variance, ceteris paribus, it would reject $\beta_3 = 0$ less often than the simpler model that I estimate, which has higher power. Because it is desirable to detect a wrong null with high probability if $\beta_3 \neq 0$, the single regression model that I estimate ( 19) is preferable to SUR.

Table 10: Empirical test of identification in the presence of school×cohort fixed effects

| | (1) Math Predicted Fixed Effects | (2) Spanish Predicted Fixed Effects |
|---|---|---|
| $\beta_1$ | -0.310 | 0.348 |
| | (0.276) | (0.277) |
| $\beta_2$ | 1.428*** | -0.352*** |
| | (0.0475) | (0.0485) |
| $\beta_3$ | 0.471 | -0.143 |
| | (0.344) | (0.348) |
| $\beta_0$ | -2.284*** | -1.269*** |
| | (0.173) | (0.179) |
| School and student controls | Yes | Yes |
| Observations | 1,810 | 1,778 |

Standard errors in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Controls: school level averages of father's education and income, teaching experience of (resp.) Math and Spanish teacher, and MSK-Intensity in the school town.