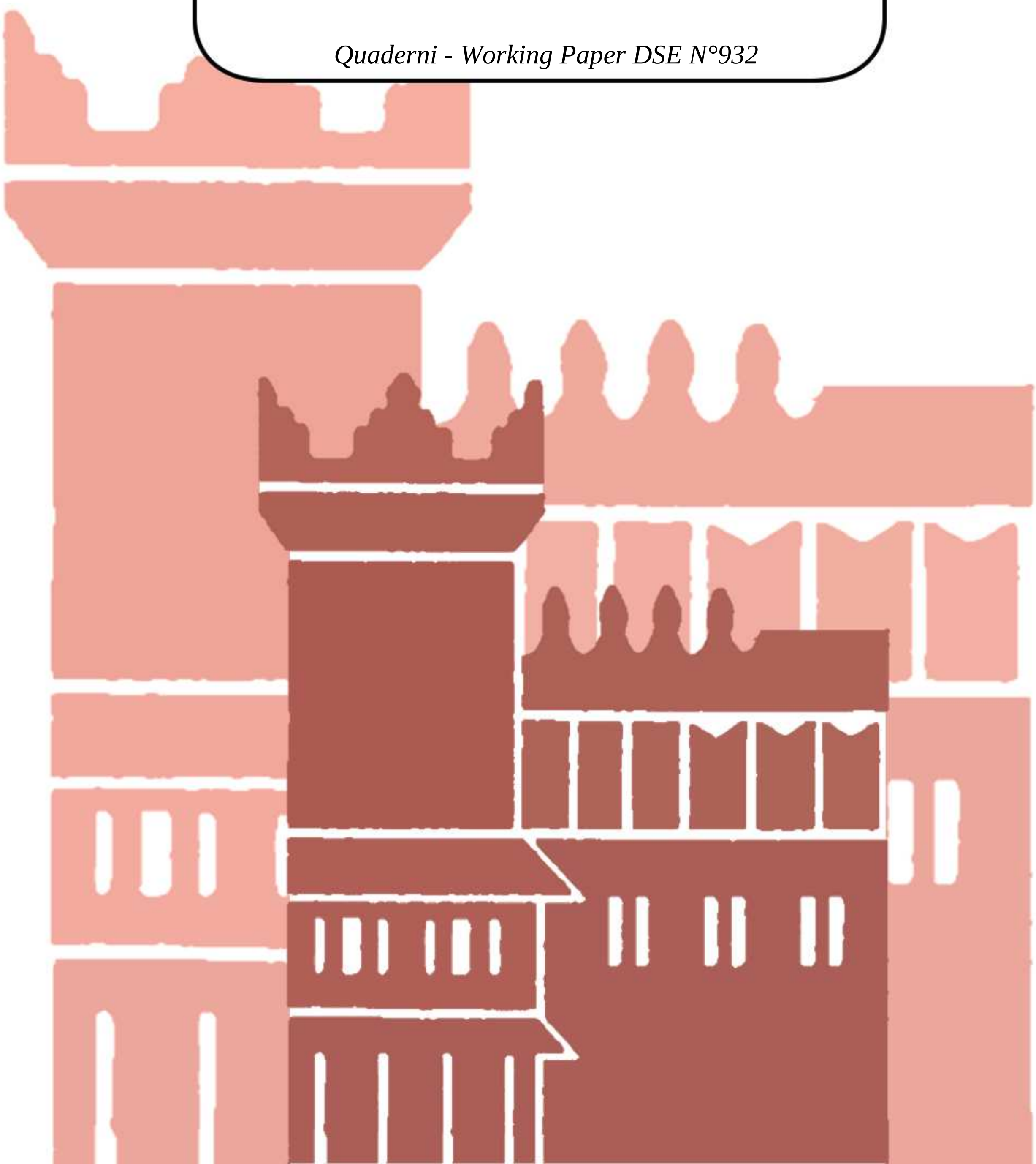
ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Separating Gender Composition Effect from Peer Effects in Education

Babak Jahanshahi

Quaderni - Working Paper DSE N°932



Separating Gender Composition Effect from Peer Effects in Education

Babak Jahanshahi*

Department of Economics, University of Bologna

Abstract This paper aims to highlight the importance of considering endogenous peer effects, as defined by Manski (1993), in order to identify gender composition effect on education outcome appropriately. Using Manski (1993) linear-in-means model, this paper illustrates that the gender composition effect that is currently estimated in education function is the function of three parameters: social multiplier, gender differences in outcome and gender composition effect (known as a gender peer effect). The appropriate gender peer effect is identified after using Graham's variance restriction method to identify and rule out a social multiplier effect. The findings suggest that a social multiplier plays a crucial role in learning process for Italian secondary and US primary students, although a gender peer effect is not as important as highlighted in previous literatures (Hoxby, 2000; Whitmore, 2005; Lavy and Schlosser, 2011) .

Keywords: Social interaction, social multiplier, gender peer effect, INVALSI, Project STAR.

JEL Classification Numbers: I21, J16

*Address: Department of Economics. University of Bologna. Piazza Scaravilli 2, 40126 Italy. E-mail: *babak.jahanshahi2@unibo.it*

I Introduction

Pupils attending school may develop their skills and abilities by receiving inputs coming from a variety of sources: teachers, school facilities, parental investments, environment and neighborhood, as well as their peers at school. The relationship between peers' interaction at school and educational outcome has attracted researchers interest since the Coleman report (Coleman et al., 1966), which was the first empirical study on peer effects at school. Subsequently, a large and multidisciplinary literature has focused on pupil's schoolmates background characteristics and abilities and their achievement at school. Several years after Manski (1993) formally discussed the difficulties in the identification of social interaction, which are potentially relevant to the study of the peer effect in education (Epple and Romano, 2011). In his seminal paper, Manski (1993) expressed three hypotheses [1] that are often used to explain the conformity of individual behavior with that of the group to which they belong. He pointed to the simultaneity problem that arises when there are both endogenous and exogenous social interactions.

Since Manski (1993), the identification of social interaction among schoolmates, commonly referred to as peer effects, has emerged as a controversial topic among socio-economic scholars. On one hand, theoretical researchers have proposed methods for the identification of social interaction (Graham, 2008; Brock and Durlauf, 2001); on the other hand, the empirical scholars (Zimmerman, 2003; Kremer and Levy, 2008; Hoxby, 2000; Whitmore, 2005; Lavy and Schlosser, 2011; Angrist and Lang, 2004; Ammermueller and Pischke, 2006; Vigdor and Nechyba, 2004; Graham, 2008) have employed either experimental or quasi-experimental research design to determine peer effect.

Only few empirical studies focus on social interaction among schoolmates of a different gender, referred to as gender peer effects and is commonly proxied by gender composition

[1] He separated peer effect to three parts as following: endogenous effect is the propensity of individual to behave in some ways varies with the prevalence of that behavior in the individual's group, exogenous effect is the propensity of individual to behave in some way varies with the characteristics of the individual's group and correlated effect is when individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments (Epple and Romano, 2011).

effect (Hoxby, 2000; Whitmore, 2005; Kang, 2007; Lavy and Schlosser, 2011). One such study is that of Hoxby (2000), who identifies idiosyncratic variation in the number of girls and achievement of students by comparing adjacent cohorts' gender and racial groups' shares. She estimated gender and race peer effects in Texas elementary schools, finding that boys and girls have higher test scores when classrooms have a larger number of female students.

Whitmore (2005) studies the share of female students on academic achievements; however, unlike Hoxby (2000) her findings are mixed (positive in kindergarten and second grade, zero in first grade and negative in third grade). In her studies Tennessees Project STAR's randomized experiment in which gender variation generated by the random assignment of students into classrooms is exploited.

Most recently, Lavy and Schlosser (2011) estimated the effects of classroom gender composition on the scholastic achievements of boys and girls in Israeli primary, middle and high schools. Following Hoxby (2000), the authors relied on idiosyncratic variations in the proportion of female students across adjacent cohorts within the same school. They found that the proportion of girls in a class has a positive and significant effect on the academic achievements of both girls and boys in high school, with the size of the estimated effects being similar for both genders. Furthermore, their exploration of the gender peer effect mechanism indicates that a higher proportion of females in a class lead to a better classroom and learning environment.

My study contributes to different strands of literature. First, it supplements existing literature on the identification of a gender peer effect (Hoxby, 2000; Whitmore, 2005; Lavy and Schlosser, 2011). However, my approach departs from other literature mentioned above by distinguishing between the gender peer effect (i.e. the variable that was aimed to be identified initially) and other determinants of the gender composition's coefficient in regression function (i.e. the gender differences in outcome and social multiplier)[2]. Hoxby (2000) claims that "when the groups are males and females (unlike racial group), there is no neat test

[2] Other determinants can be derived from applying Manski's linear-in-means model to the gender peer effect framework, as further described in Chapter 2.

of whether a group's peer effects all operate through peer achievement". Within Manski's framework, this means one cannot separate the exogenous effect of having more females in the classroom from the fact that females might be better peers and have higher scores. Lavy and Schlosser (2011) do not consider the spillover effects of pupils achievements in investigating the overall payoff from all the possible mechanisms through which gender peer effects might be at play; instead, their analysis is limited to the few channels through which a gender peer effect might be at work. Whitmore (2005) mentions that having a predominately female class in the second grade substantially increases students test scores, which can only be partly explained by being exposed to higher quality peers (as girls' scores are higher than those of boys). Therefore, she claims that there should be something further about having a predominately female class per se, although her study does not precisely distinguish between different possible effects.

Second, I contribute to the parts of the literature on social interaction that aim to overcome the reflection problem [3] in order to estimate the effects of the endogenous social multiplier in a quasi-experimental framework. Finally, to my knowledge this is the first paper to estimate a social multiplier in an Italian school.

The remainder of this paper is organized as follows. Section 2 presents the theoretical concept to show the problem of identification based on Manski's linear-in-means model. Section 3 suggests the empirical strategies to solve the identification problems, while section 4 presents the data set. Section 5 presents the results, before section 6 summarizes the findings and provides a brief conclusion.

II Gender peer effect in linear-in-means form

In order to show the identification problem of estimating a gender peer effect, I assume that a social interaction takes the linear-in-means form as in Manski (1993). Assume:

[3] The term "reflection problem" is used to characterize the simultaneity problem that arises when there are both endogenous and exogenous social interaction Epple and Romano (2011)

$$y_{ci} = \alpha_0 + \alpha_1 x_i + \alpha_2 x_c + \beta \bar{y}_c + \epsilon_{ci} \tag{1}$$

Where; In each grade, denote classes with c and individual with i. y is individual achievement in school, x_i is a dummy variable denoting the gender of individual i, which is equal to 1 if i is a girl. x_c is the proportion of girls in each class (i.e. $E(x_i|c)$), $\bar{y}_c$ is the average achievement of individual i in the class, ϵ_{ci} are unobserved attributes that directly affect y. Following Manski (1993), I assume $E(\epsilon_{ci}|c, x_i) = c'\sigma$, which captures the correlated effect.

One should note the two important restrictions associated with the specification introduced above. Firstly, it is implicitly assumed that the gender composition effect is identical across gender, and secondly, the endogenous effect is homogenous across gender, meaning that the average achievement of an individual affects all the students identically, regardless of their gender. The first assumption is verified by looking at the results of previous literature in the field (Whitmore, 2005; Lavy and Schlosser, 2011). Both Hoxby (2000) and Lavy and Schlosser (2011) findings show that the proportion of girls in the classroom affect both genders virtually identical, while the second assumption is logical given that the gender composition effect is internalized in the model as an exogenous peer effect.

Average achievement within a class leads to:

$$\bar{y}_c = \alpha_0/(1-\beta) + 1/(1-\beta)(\alpha_1 + \alpha_2)x_c + 1/(1-\beta)c'\sigma \tag{2}$$

A reduced form is obtained by replacing Eq. (2) into Eq. (1):

$$y_{ci} = \gamma\alpha_0 + \alpha_1 x_i + ((\gamma - 1)\alpha_1 + \gamma\alpha_2)x_c + \gamma c'\sigma \tag{3}$$

Where; $\gamma = 1/(1-\beta)$ is a social multiplier, namely the ratio between the average cumulative response and the individual response following an exogenous shock. From Eq. (3), one can clearly see the identification problem that arises in the study of peer effect, as discussed by Manski (1993): by OLS regression of individual achievement on gender composition in the classroom, only the composite parameters $\alpha_0\gamma$, α_1, $((\gamma-1)\alpha_1+\gamma\alpha_2)$ and $\gamma\sigma$ are identified. Moreover, identification of the composite parameters does not enable us to distinguish between the two social effects (endogenous and exogenous ones). As one can see from Eq. (3), based on Manski's linear-in-means model the coefficient that are estimated so far by regressing gender composition on educational outcome (i.e $((\gamma-1)\alpha_1+\gamma\alpha_2)$) is formed by three separate elements: the effect of having more girls in the classroom (α_2) , the difference between girls and boys in educational outcome (α_1) and the social multiplier (γ).

III Empirical Strategy

In order to solve the identification problem mentioned in chapter 2, the social multiplier (γ) is estimated first, which allows driving a gender peer effect that is solely due to the existence of more girls in the class (α_2) by estimating Eq. (3).

A Identification of Social Multiplier

Graham (2008) proposed a method for the identification of a social multiplier (γ in equation 3), by exploiting differences in variances across groups. For a linear form of social interaction, he defined the unconditional between-group variance of means outcome as the sum of the variance of any group level heterogeneity (classroom certain characteristics such as teacher quality), between-group variance of any individual-level heterogeneity (variability in average student ability) and the strength of any social interaction (peer effect). Therefore, in the presence of social interaction, between-group variation in outcome should reflex between-group variation in 'peer quality'. Following Galbiati and Zanella (2012), we can rewrite the reduced form model from equations (2) and (3) in variance components. The transformation of group-level heterogeneity (α_c=α_2 girls + $\sigma c'$), individual-level heterogeneity (ϵ_{ci}=α_1 gender) and the group level average of individual-level heterogeneity ($\bar{\epsilon_c}$= $\alpha_1 girls$) yields the following behavioral equations:

$$y_{ci} = \gamma\alpha_c + \epsilon_{ci} + (\gamma - 1)\bar{\epsilon}_c \tag{4}$$

$$\bar{y_c} = \gamma(\alpha_c + \bar{\epsilon_c}) \tag{5}$$

Graham (2008) proved that under some specific assumptions discussed below, γ^2 can be identified by using the following conditional and unconditional restrictions:

$$E[G_c^b - \theta W_{2c} - \gamma^2 G_c^w | W_{1c}, W_{2c}] = 0 \tag{6}$$

$$E[\begin{pmatrix} W_{1c} \\ W_{2c} \end{pmatrix}(G_c^b - \theta W_{2c} - \gamma^2 G_c^w)] = 0 \tag{7}$$

Where; W_{1c} and W_{2c} are two vectors containing observable classroom-level information, W_{1c} denotes class size (small vs. large) and W_{2c} denotes other classroom-level information such as the share of educated parents, share of immigrants in the classroom, etc. G_c^w and G_c^b are within- and between- group statistics, respectively. (For more details, see supplement part of Graham (2008) and Galbiati and Zanella (2012)).

Eq.(7) delivers the appropriate specification to estimate (i.e. by GMM) the social multiplier, γ^2, using W_{1c} as an instrumental variable.

The three primitive assumptions that guarantee identification are as follows:

- Independent Random Assignment: Teacher and students assignment to classroom must be random.

- Stochastic Separability: The population variance of small and large classroom teacher effectiveness must be the same.

- Peer Quality Variation: This is a rank restriction, which requires that the variance of peer quality differs between the two types of classrooms.

B Identification of Composite Parameters

The model based on Eq. (3) suggests that regression of "gender composition" on educational outcome delivers the coefficient of the following form:

$$\delta = (\gamma - 1)\alpha_1 + \gamma\alpha_2 \tag{8}$$

δ is estimated for two case studies, namely the US and Italy. The first case study is based on a randomized experiment, while, for the second case study, idiosyncratic variation in gender composition across adjacent cohort is employed in order to gain a clean estimate of δ.

C Identification of Gender Peer Effect

In order to recover a gender peer effect and its standard deviation, a bootstrapping method is used to approximate the distribution of a statistic by a Monte Carlo simulation.

IV Data

The empirical analysis is based on two case studies: elementary school students in the US and secondary students in Italy. The reasons for including two different case studies are threefold. Firstly, in order to investigate the gender peer effect in both primary and secondary schools. Secondly, in order to gain a better understanding of the importance of endogenous effects by comparing my results with those from Hoxby (2000) and Whitmore

(2005), two main contributions to existing literature on gender peer effect. And finally, the Italian case study is very applicable in order to introduce a method for investigating social multiplier in a non-experimental framework.

A US

The assessment of gender peer effect in the learning process is conducted by using data from the class size reduction experiment Project STAR. According to Word et al. (1990), Project STAR was started in the fall of 1985, whereby kindergarten students were randomly assigned to one of three class types within their school: small, regular and regular with a full-time teacher's aide. Thereafter, teachers were randomly assigned to one of these three class types.

The within-school randomization was implemented in 79 schools and ultimately included 11,600 students. In the experiment, a single cohort of children was assigned to small or regular classes from kindergarten through to third grade, before all students returned to regular sized classes in fourth grade.

B Italy

For the Italian primary students, the data requirements are fulfilled by the INVALSI data set for the universe of Italian primary and secondary schools in the academic years 2009-10 and 2010-2011. INVALSI (the National Institute for the Evaluation of the Education System) is in charge of designing and administering standardized education tests in Italy. Since 2008, the tests have been administered on an annual basis.

The recent waves of this data set collected data for the population of primary and lower secondary students in their second, fifth, sixth and eighth Italian grades. For each student, the data set contains information on class size and grade in the school, immigrant status based on citizenship and language spoken at home, test scores in Italian and Math, gender, age and family background information.

Tables 1, 2 and 3 present the number of observations, the mean, the standard deviation, and the minimum and the maximum values of math and reading scores of boys, girls and the overall population for second, fifth and eighth graders, respectively. For example, the fifth graders standardized reading test had a mean of around 0.7 points and a standard deviation of around 0.17 points in 2009-2010. The average female scored 0.02 points – around a 0.12 standard deviation – higher than the average male.

V Results

A Social Multiplier and Gender Peer Effect in US Primary Schools

Full details on the validity of identification assumptions one need to identify social multiplier with experiment Project STAR are provided by Graham (2008). However, he limited his analysis to kindergarten students. Table 4 reports Graham (2008) findings for kindergarten students as well as the social multipliers that I assess for second and third graders using Tennessee's Project STAR experiment. The estimations of social multipliers for second graders are 2.23 and 2.14 for math and reading, respectively, and the standard errors of parameter recovered by using the delta method. These are almost the same as estimated for kindergarten students. Third graders' social multipliers are 1.5 and 2 for math and reading, respectively, which suggests that a social multiplier might be less determinant for upper graders. The first graders are ruled out from the analysis, given that, according to Whitmore (2005), kindergarten was not required in Tennessee at the time of Project STAR, and consequently there was a large influx of new entrants in first grade of significantly lower quality than kindergarten entrants who might have disrupted classrooms.

The estimations of gender peer effects are presented in Table 5. After accounting for the roles of a social multiplier and the differences between gender in outcome, gender peer effects lost most of their initial magnitude. However, one should note that, in the cases where social multipliers are not significantly different from one, social interactions are not at place, and the female share coefficient (δ) reflects the gender peer effect coefficient (α_2).

It is important to highlight that a bootstrapping method is used to approximate the distribution of a statistic by a Monte Carlo simulation in order to recover the gender peer effect and its standard deviation.

B Social Multiplier and Gender Peer Effect in Italy Primary and Secondary Schools

In order to identify the social multiplier among Italian students, the discontinuity in the relationship between enrollment and class size at an enrollment multiple of 25, which is induced by the so-called "Maimonides' rule"[4], is employed. This discontinuity induced classes of different sizes, prompting the need to employ Grahams method. Tables 6 and 7 and panel B of table 3 show descriptive statistics for the grades with enrollments in a range close to the points of discontinuity. These are the grades with enrollment in the set of intervals $\{[22, 30], [47, 55], [72, 80], [97, 105], [122, 130], [147, 155], [172, 180], [197, 205], [222, 230], [247, 255], [272, 280], [297, 305], [322, 330], [347, 355], [372, 380], [397, 404]\}$. Around 17 percent of the total grades are in these intervals after accounting for a +10% margin of flexibility [5]. As is shown in the tables, the average characteristics of classes in the discontinuity sample are remarkably similar to those for the full sample.

B.1 Assumptions Verification

In this section, I assess the three required assumptions in order to identify social multiplier: peer quality variation, independent random assignment and stochastic separability. The approach adopted here is based on non-experimental methods in evaluation research (Campbell, 1969): regression discontinuity design. This method utilizes verifying the neces-

[4] This term was first used by Angrist and Lavy (1999).

[5] For example, for the first interval enrollment that contains 25 and 26 students is excluded

sary assumptions in order to estimate social multiplier appropriately.

1. Peer Quality Variation The idea of using RDD to identify class size effect comes from what Angrist and Lavy (1999) termed Maimonides' rule, in which they exploit the fact that class size is partly determined by a known discontinuity function of observed covariates (enrollment in a grade). For my purpose, the importance of Maimonides rule is that it has been used to determine the division of enrollment grades into classes in Italian public schools. Based on Italian law, class size cannot be larger than 25, with a margin of flexibility of +10 percent. Moreover, it cannot be smaller than 10, with a margin of flexibility of -10%. Let Z be the total enrollment in a grade and C the number of classes; subsequently, the rule for class size disregarding the margins of flexibility is:

$$\bar{S} = \frac{Z}{Int(\frac{Z-1}{25}) + 1} \tag{9}$$

   Where $lnt(x)$ is the largest integer smaller or equal to x. Based on equation (14), the theoretical class size is a function of grade (in a particular school) enrollment, which displays discontinuities at multiples of 25. We can see the predicted and actual class size in Italian elementary school in figure 1 (taken from Ballatore et al. (2012)).

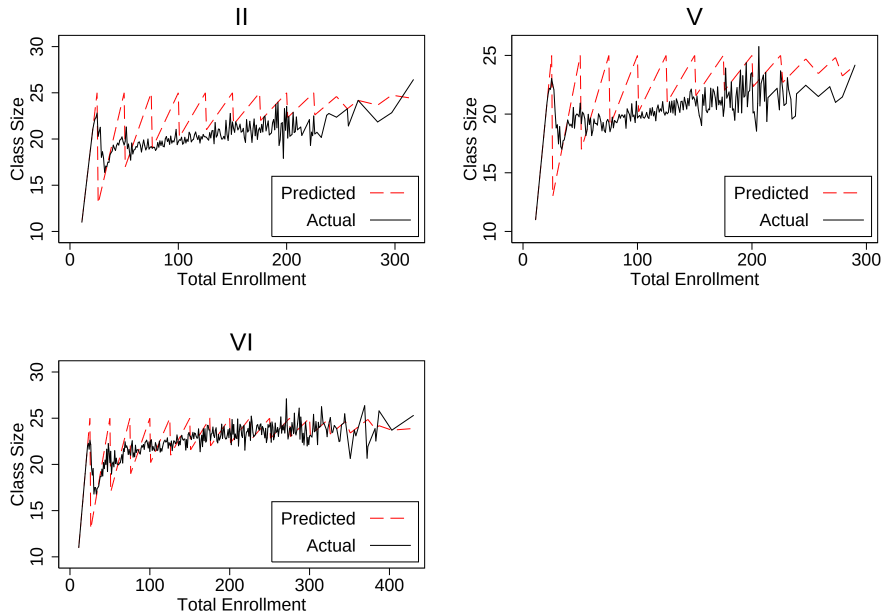
   On the left of each threshold, the theoretical class size is larger than on the right, with this feature of the rule offering a source of variation in peer equality. As I will show in the next section, the variance of peer quality indeed differs between two types of the classroom. Therefore, one of the three assumptions is required for identification to be verified (i.e. rank condition is satisfied).

2. Independent Random Assignment [6]

   The attractive feature of RDD is the fact that it allows testing the validity of its

[6] this assumption is also called double randomization assumption, which means students and teachers should independently and randomly assigned to the classroom

Figure 1: Predicted and actual class size in Italy



Note: Each graph shows the predicted (red line) and actual (black line) class size in different grades

identification condition, which is parallel to the assumption of independent random assignment. The condition for identification based on RDD requires that no discontinuity takes place at the threshold for selection in the counterfactual world. This is called the orthogonality condition, which is as follows:

$$(Y1, Y0) \perp I|S = s \tag{10}$$

Where (Y_1, Y_0) are the two potential outcomes. I is the binary variable that denotes treatment status, with $I = 1$ for small classroom and $I = 0$ for larger ones. Treatment status depends on an observable unit characteristic S (enrollment), and there exists a known point in the support of S where the probability of participation changes discontinuously (enrollment equal to 25).

Tables 8 and 9 present the test for this assumption based on the idea of comparing units marginally above and below the threshold with respect to variables whereby:

- cannot be affected by the treatment;
- are affected by the same unobservable that is relevant for the outcome.

With few exceptions, the evidences in tables 8 and 9 suggest that the existence of discontinuities in pre-treatment variables is unlikely to be correlated with potential outcomes. To confirm this result and ensure that the exceptions in tables 8 and 9 are only a spurious correlations, table 10 indicates the Pearson's chi-squared Test for the random assignment of girls in the classroom. This test was first used by Ammermueller and Pischke (2006). The results of the test suggest that girls are randomly spread across the classes of different size, which provides further evidence in support of an "Independent Random Assignment" (for eighth graders, only a Pearson's chi- squared Test for all the country is measurable due to limitations in the data set).

The evidences in tables 8, 9 and 10 allow one to reject the presence of discontinuities in pre-treatment variables that are likely to be correlated with potential outcomes. In other words, the schools below and above the threshold are comparable.

3. Stochastic separability

This assumption states that the teacher effectiveness variation across two types of classroom must be equal, and is not valid if the distribution of teacher characteristics is not similar across classrooms of different sizes. As we compare classes with different size across different schools, it is very unlikely that teachers are sorted across classes. However, to further test this assumption, a sensitivity analysis test suggested by Graham (2008) is performed. The results of the sensitivity analysis test suggest that the typical difference in effectiveness across a pair of teachers would have to be implausibly large in small versus large classrooms to produce social multiplier estimates of the size reported in table 11, if, in fact, there were no peer effects. (For details of sensitivity analysis, see supplement to Graham (2008))

B.2 Results

Table 11 reports the estimate of γ^2 using 2009-2010 wave of INVALSI dataset for the second, fifth and eighth graders by estimating equation 7. The first, third and fifth columns report the results for math and the second, fourth and sixth for reading (Italian). The estimates of a social multiplier are 2.8, 1.88 and 3.24 for math and 1.52, 3.03 and 3.85 for reading in the second, fifth and eighth grades, respectively. These findings suggest that social interaction plays an important role in the learning process. In contrast to the fifth and eighth graders, the null hypothesis that $\gamma^2 = 1$ is not rejected at the 90% confidence level for second graders; therefore, one cannot reject the hypothesis of no peer interaction for second graders.

Panel B of table 11 shows the first stage results of the estimate. The coefficient of variable "small" is statistically significant, which supports the first assumption of peer quality variation. The first stage F- statistics is large, suggesting that the instrument is not weak. In order to check the robustness of the results, table 12 presents the social multiplier calculated for the first two thresholds (enrollments less than 60), with the results proving robust across the two different samples.

Following the empirical method employed by Hoxby (2000) and Lavy and Schlosser (2011), gender peer effects for Italian 8th graders (Eq. (3)) are estimated by relying on idiosyncratic variation in gender composition across adjacent cohorts within the same grade in the same school (eighth grade is the only grade whereby one can match cohorts of adjacent years by using the Invalsi data set). This approach proposes a persuasive solution for the two possible sources of confounding factors: self-selection of students into the schools and correlation between school characteristics and gender composition.

The estimations of gender peer effects for Italian students are presented in Table 13. After considering the role of a social multiplier and the differences between genders in terms of outcome, the gender peer effect is relatively large and negatively significant in math and approximately zero and not significant in reading. This is consistent with the findings from Whitmore (2005)empirical study indicating that a peer effect in school deteriorates educational outcomes for upper grade females.

VI Conclusion

In this paper, I empirically measure the extent of gender peer effects in Italian secondary and US primary schools on students academic achievements. Using Manski (1993) linear-in-means model, I was able to disentangle two different mechanisms through which a higher proportion of females in the class might affect students academic achievements: a social multiplier and a gender composition effect. It is shown that the two mentioned mechanisms, along with gender differences in outcome, form the gender composition coefficient estimated to date by researchers in order to find gender peer effect in school on academic achievement.

The project STAR experiment allows identifying a gender peer effect for US primary students, while this is identified for Italian secondary students by using idiosyncratic variation in gender composition across an adjacent cohort within the same school. In order to disentangle the multiplier's effect Graham (2008) conditional variance restriction method is employed.

With one exception, the evidence provided in this paper suggests that a social interaction plays a crucial role in the learning process for primary pupils in the US and secondary pupils in Italy. However, the gender composition effect is not as important as previously thought, after accounting for a social multiplier and gender gap in the outcome. The general implication of these findings is that in contrast to gender mix of class, the spillover effects of pupils achievements should be taken into account in inter- and intra-school resource allocation in elementary schools. Furthermore, findings show that higher proportions of females in the math classroom deteriorate the educational outcome of upper grade male pupils. Indeed, this is consistent with the findings from Whitmore (2005) empirical study.

This study does not control for a heterogeneous social multiplier effect across gender and is unable to rule out the possibility that the female proportion in the classroom might differ in importance for education outcome between the two genders. However, the results provide important insight towards understanding the relative role of a social multiplier and gender composition effect.

References

Ammermueller, A. and J.-S. Pischke (2006). Peer effects in european primary schools: Evidence from pirls. Technical report, National bureau of economic research.

Angrist, J. D. and K. Lang (2004). Does school integration generate peer effects? evidence from boston's metco program. *The American Economic Review 94*(5), 1613–1634.

Angrist, J. D. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics 114*(2), 533–575.

Arcidiacono, P. and S. Nicholson (2005). Peer effects in medical school. *Journal of Public Economics 89*(2), 327–350.

Ballatore, R., M. Fort, and A. Ichino (2012). Babels tower in the classroom?

Brock, W. A. and S. N. Durlauf (2001). Discrete choice with social interactions. *The Review of Economic Studies 68*(2), 235–260.

Brock, W. A. and S. N. Durlauf (2007). Identification of binary choice models with social interactions. *Journal of Econometrics 140*(1), 52–75.

Campbell, D. T. (1969). Reforms as experiments. *American psychologist 24*(4), 409.

Coleman, J. S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld, and R. York (1966). Equality of educational opportunity. *Washington, dc.*

Cook, T. D., D. T. Campbell, and A. Day (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Houghton Mifflin Boston.

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.

Falk, A. and A. Ichino (2006). Clean evidence on peer effects. *Journal of Labor Economics 24*(1), 39–57.

Galbiati, R. and G. Zanella (2012). The tax evasion social multiplier: Evidence from italy. *Journal of Public Economics 96*(5), 485–494.

Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica 76*(3), 643–660.

Harwarth, I., E. DeBra, and M. Maline (1997). *Women's Colleges in the United States: History, Issues And, Challenges.* DIANE Publishing.

Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research 43*(6), 387–425.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research.

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.

Lavy, V. and A. Schlosser (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics 3*(2), 1–33.

Mael, F., A. Alonso, D. Gibson, K. Rogers, and M. Smith (2005). Single-sex versus coeducational schooling: A systematic review. doc# 2005-01. *US Department of Education*.

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

Morse, S. (1998). *Separated by sex: A critical look at single-sex education for girls.* Amer Assn of Univ Women.

Vigdor, J. and T. Nechyba (2004). Peer effects in elementary school: Learning from apparentrandom assignment. Technical report, Working Paper.

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

Word, E. R. et al. (1990). The state of tennessee's student/teacher achievement ratio (star) project: Technical report (1985-1990).

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

Tables:

Table 1: Discriptive Statistics - grade 2

	mean	sd	min	p25	p50	p75	max	var
A. 2009-2010								
Grade 2: 5969 schools,22745 classes								
All score math	.62	.2	0	.46	.6	.78	1	.04
All score ita	.66	.23	0	.5	.69	.85	1	.053
boy score math	.63	.2	0	.46	.61	.79	1	.04
boy score ita	.64	.23	0	.46	.69	.85	1	.054
girl score math	.62	.2	0	.46	.61	.79	1	.04
girl score ita	.67	.23	0	.5	.73	.85	1	.051
class size	20.6	3.6	11	18	21	23	35	13
B.2010-2011								
Grade 2: 7337 schools, 26628 classes								
All score math	.66	.19	0	.53	.68	.78	1	.037
All score ita	.72	.19	0	.6	.76	.86	1	.035
boy score math	.66	.19	0	.53	.68	.82	1	.037
boy score ita	.71	.19	0	.6	.76	.87	1	.035
girl score math	.65	.19	0	.53	.64	.78	1	.037
girl score ita	.73	.18	0	.63	.76	.87	1	.03
class size	19	3.8	11	17	20	22	35	14.7

Source: Invalsi data for academic years 2009-10 and 2010-2011. Performance in a test is measured as the fraction of correct answers.

Table 2: Discriptive Statistics - grade 5

	mean	sd	min	p25	p50	p75	max	var
A. 2009-2010								
Grade 5: 5937 schools, 22846 class								
All score math	.65	.18	0	.52	.66	.79	1	.034
All score ita	.7	.17	0	.59	.74	.84	1	.03
boy score math	.66	.19	0	.52	.68	.82	1	.035
boy score ita	.69	.18	0	.58	.72	.83	1	.031
girl score math	.64	.18	0	.5	.64	.77	1	.033
girl score ita	.71	.17	0	.61	.74	.84	1	.03
class size	20.8	3.7	11	18	21	24	35	13.8
B. 2010-2011								
Grade 5: 7374 schools, 27303 classes								
All score math	0.69	.17	0	.59	.72	.83	1	.028
All score ita	.74	.14	0	.65	.75	.85	1	.02
boy score math	.7	.16	0	.59	.72	.83	1	.02
boy score ita	.73	.15	0	.65	.75	.84	1	.02
girl score math	.69	.17	0	.57	.7	.83	1	.028
girl score ita	.74	.14	0	.65	.77	.85	1	.02
class size	19	3.8	11	17	19	22	35	14.4

Source: Invalsi data for academic years 2009-10 and 2010-2011. Performance in a test is measured as the fraction of correct answers.

Table 3: Discriptive Statistics - grade 8

	mean	sd	min	p25	p50	p75	max	var
A. 2009-2010 full sample								
Grade 8: 3760 schools, 21577 classes								
All score math	.49	.18	0	.36	.47	.62	1	.03
All score ita	.68	.17	0	.57	.71	.81	1	.03
boy score math	.51	.19	0	.38	.5	.66	1	.035
boy score ita	.65	.18	0	.55	.69	.79	1	.033
girl score math	.46	.17	0	.34	.45	.58	1	.029
girl score ita	.7	.16	0	.61	.73	.82	1	.026
class size	20	4.3	11	17	21	24	33	18.7
B. 2009-2010 discontinuity sample								
Grade 5: 613 schools, 3604 classes								
All score math	.48	.18	0	.36	.47	.62	1	.03
All score ita	.68	.18	0	.57	.71	.81	1	.032
boy score math	.51	.19	0	.36	.5	.64	1	.035
boy score ita	.65	.19	0	.54	.67	.79	1	.035
girl score math	.46	.17	0	.34	.45	.58	1	.03
girl score ita	.7	.16	0	.61	.74	.82	1	.027
class size	20	4.4	11	17	21	24	32	19.3

Source: Invalsi data for academic years 2009-10 and 2010-2011. Performance in a test is measured as the fraction of correct answers.

Table 4: Social Multiplier - Tenessee project

	Kindergarten Math	Re	2nd grade Math	Re	3rd grade Math	Re
G_c^w (coefficient: γ^2)	3.47	5.28	5	4.58	2.26	4.05
	(1.03)	(2.48)	(1.8)	(2.1)	(1.3)	(1.04)
Social multiplier (γ)	1.86***	2.3***	2.23***	2.14***	1.5***	2.01***
(delta method)	(0.27)	(0.54)	(0.4)	(0.49)	(0.44)	(0.26)
p-value H_0*:* $\gamma^2 = 1$	0.018	0.086	0.02	0.09	0.34	0.004
B: *First stage*						
F-stat.	46.8	19.0	57.08	38.88	45.58	56.7
Number of classroom	317	317	331	331	330	325
School fixed effects	✓	✓	✓	✓	✓	✓

Notes. Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Gender peer effect - US

	Kindergarten Math	Re	2nd grade Math	Re	3rd grade Math	Re
Female Share(δ_1)	0.42**	0.35**	0.24	0.503**	-0.303	-0.33
	(0.186)	(0.17)	(0.281)	(0.250)	(0.252)	(0.26)
Gender peer effect*	0.17***	0.066***	0.09***	0.13***	-0.2***	-0.27***
	(0.003)	(0.002)	(0.004)	(0.004)	(0.005)	(0.003)
Control						
School fixed effects	✓	✓	✓	✓	✓	✓
Classroom type	✓	✓	✓	✓	✓	✓
Socio-economic statues	✓	✓	✓	✓	✓	✓
race	✓	✓	✓	✓	✓	✓
Observations	5707	5629	5723	5731	5829	5751
R-squared	0.264	0.263	0.256	0.254	0.228	0.2

Notes. Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.
* a bootstrapping method is utilized to approximate the distribution of a statistic by a Monte Carlo simulation.

Table 6: Discriptive Statistics - discontinuity sample grade 2

	mean	sd	min	p25	p50	p75	max	var
A. 2009-2010								
Grade 2: 986 schools, 3632 classes								
All score math	.62	.2	0	.46	.6	.78	1	.04
All score ita	.65	.23	0	.5	.69	.85	1	.05
boy score math	.62	.2	0	.46	.61	.78	1	.04
boy score ita	.64	.23	0	.46	.6	.78	1	.04
girl score math	.62	.2	0	.46	.6	.75	1	.04
girl score ita	.67	.23	0	.5	.73	.85	1	.05
class size	21	3.8	11	18	21	24	32	15
B. 2010-2011								
Grade 2: 1162 schools, 4228 classes								
All score math	.66	.19	0	.53	.68	.82	1	.04
All score ita	.72	.19	0	.6	.76	.87	1	.035
boy score math	.66	.19	0	.53	.68	.82	1	.038
boy score ita	.72	.19	0	.6	.76	.87	1	.035
girl score math	.65	.19	0	.53	.68	.78	1	.037
girl score ita	.73	.18	0	.63	.76	.87	1	.03
class size	19.6	4	11	17	20	23	30	16

Source: Invalsi data for academic years 2009-10 and 2010-2011. Performance in a test is measured as the fraction of correct answers.

Table 7: Discriptive Statistics - discontinuity sample grade 5

	mean	sd	min	p25	p50	p75	max	var
A. 2009-2010								
Grade 5: 1021 schools, 3791 classes								
All score math	.65	.18	0	.52	.66	.79	1	.34
All score ita	.7	.17	0	.59	.74	.84	1	.03
boy score math	.66	.19	0	.52	.68	.82	1	.35
boy score ita	.69	.17	0	.58	.72	.84	1	.03
girl score math	.64	.18	0	.5	.64	.79	1	.03
girl score ita	.71	.17	0	.6	.75	.84	1	.03
class size	21	3.9	11	18	21	24	29	15.3
B. 2010-2011								
Grade 5: 1185 schools, 4371 classes								
All score math	.7	.17	0	.59	.72	.83	1	.028
All score ita	.74	.14	0	.65	.77	.85	1	.02
boy score math	.71	.17	0	.59	.72	.82	1	.03
boy score ita	.74	.14	0	.65	.75	.85	1	.02
girl score math	.69	.17	0	.59	.72	.83	1	.028
girl score ita	.75	.14	0	.67	.77	.85	1	.019
class size	19	4	11	17	19	23	29	16

Source: Invalsi data for academic years 2009-10 and 2010-2011. Performance in a test is measured as the fraction of correct answers.

Table 8: Random Allocation Test - discontinuity sample grade 2

VARIABLES	Share with high educated parents	Share with low skilled parents	Share with num of imigrants	share of girls
gap at the threshold	-0.008***	0.001*	-0.001	-0.006*
	(0.001)	(0.001)	(0.001)	(0.0035)
Observations	69067	69067	69067	68417
R-squared	0.002	0	0	0

Notes. Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Random Allocation Test - discontinuity sample grade 5

VARIABLES	Share with high educated parents	Share with low books at home	Share with num of imigrants	share of girls
gap at the threshold	0	-0.009	0	-0.003
	(0.001)	(0.006)	(0.001)	(0.004)
Observations	73822	73822	73822	73181
R-squared	0	0	0	0

Notes. Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Pearson's chi-squared Test for random assignment of girls in the classroom

	Grade 2	Grade 5	Grade 8
All the Country			
Pearson's test statistics	1995.873	2424.859	313.6128
degree of freedom	2644	2770	2961
p-value	1	1	1
North			
Pearson's test statistics	940	928	
degree of freedom	1220	1130	
p-value	1	1	
Center			
Pearson's test statistics	372	439	
degree of freedom	477	492	
p-value	1	0.96	
South			
Pearson's test statistics	682	911	
degree of freedom	947	1148	
p-value	1	1	

Notes. The degrees of freedom are $\sum_s^S (n_{class} - 1)/J - 1$. S is the total number of schools for a given grade and J is the number of possible values taken by the characteristic one wants to test the random assignment.

Table 11: Social Multiplier - Italy

	2nd grade		5th grade		8th grade	
	Math	Re	Math	Re	Math	Re
G_c^w (coefficient: γ^2)	7.86	2.33	3.55	9.21	10.5	14.8
	(5)	(0.9)	(1.4)	(1.3)	(4.13)	(2.9)
Social multiplier (γ)	2.8***	1.52***	1.88***	3.03***	3.24***	3.85***
(delta method)	(0.89)	(0.29)	(0.38)	(0.22)	(0.63)	(0.38)
p-value H_0*:* $\gamma^2 = 1$	0.17	0.14	0.07	0	0.02	0
B: *First stage*						
F-stat.	10.44	19.11	49.93	11.36	4.1e+10	1.2e+10
p-value	0.0012	0	0	0.0008	0	0
Number of classroom	3627	3623	3791	3791	3812	3811
School fixed effects	✓	✓	✓	✓	✓	✓

Notes. Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Social Multiplier - Italy first two threshold

	2nd grade		5th grade		8th grade	
	Math	Re	Math	Re	Math	Re
G_c^w (coefficient: γ^2)	7.86	2.33	3.55	9.21	7.77	10.84
	(6.6)	(1.2)	(1.88)	(1.76)	(6.13)	(4.05)
Social multiplier (γ)	2.8***	1.52***	1.88***	3.03***	2.78***	3.29***
(delta method)	(1.19)	(0.39)	(0.5)	(0.29)	(1.1)	(0.6)
p-value H_0*:* $\gamma^2 = 1$	0.3	0.26	0.17	0	0.27	0.01
B: *First stage*						
F-stat.	5.97	10.95	27.59	6.45	1.4e+12	8.3e+10
p-value	0.015	0.001	0	0.01	0	
Number of classroom	785	788	817	817	281	282
School fixed effects	✓	✓	✓	✓	✓	✓

Notes. Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 13: Gender peer effect - Eighth Italian graders

	Math	Re
Female Share(δ)	-0.025	-0.22***
	(0.08)	(0.1)
Gender peer effect*	-0.56***	-0.003
	(0.005)	(0.003)
Control		
School fixed effects	✓	✓
Time fixed effect	✓	✓
Observations	15102	15102
R-squared	0.8	0.7

Notes. Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

* a bootstrapping method is utilized to approximate the distribution of a statistic by a Monte Carlo simulation.

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it