Social Interactions in Schooling

Article · September 2004

Source: RePEc

2 authors, including:

Alfonso Rosolia
Banca d'Italia

Social interactions in schooling*

Piero Cipollone[‡] Alfonso Rosolia[‡]

Bank of Italy
Research Department

This version: February 2003.

PRELIMINARY

Abstract

We exploit the 1980 earthquake in southern Italy and the subsequent relief from mandatory military service granted to all males in the regions hit by the seism to estimate the strength of endogenous social interactions in schooling achievements. Preliminary results point to a significant and strong effect of interactions not mediated by markets.

*The views expressed in this paper do not necessarily reflect those of the Bank of Italy.

[‡]We thank Giovanni Seri for his help with handling the Census data. Correspondence: Research Department, Bank of Italy, Via Nazionale 91, 00184 - Roma - Italy.

1 Introduction.

The last decade has seen a burgeoning literature involving concepts such as social interactions, peer effects, social norms, band-wagon effects and the like. A sense of this upsurge can be given by figure (1) where we report the number of EconLit hits found searching abstracts and titles for *social interaction* and *peer-effect*. Out of nearly 200 hits since 1973, more than a half are posterior to 1995. The theoretical underpinnings of these concepts are relatively clear, and boil down to the workings of some kind of market failure such that in equilibrium the dependence of individual behaviours on the aggregate behaviour of some reference group is not considered by agents when making their decisions (e.g. externalities). The empirical counterpart is much weaker because of serious identification problems. As made clear by the seminal work of Manski (1993), this is basically due to the fact that average behaviour, which supposedly feeds back into individual outcomes if such mechanisms are at work, carries no additional information on top of that already embodied into observed individual outocomes. In the literature this missing information is generally recovered by imposing a structure to the problem being studied, either making assumptions on the functional form for the link between aggregate and individual outcomes (e.g. a non-linear function of individual behaviours) or excluding any other possible relationship between average characteristics of a group and individual outcomes. This is a legitimate strategy when the structure derives from the theoretical predictions of a model; it is not when the assumptions are made exclusively to achieve identification. When this is the case a more credible strategy is to rely on genuine additional information beyond that provided by individual outcomes. This boils down to finding suitable instruments for aggregate outcomes.

The rewards for a reliable estimation of such effects are high, especially when it comes to policy design, due to their increasing importance as a major ingredient in several fields of economic theory such as economic growth, spatial agglomeration, inequality, technological choices and so on (see Brock and Durlauf (2000) and Moffitt (2000) for comprehensive lists).

In this paper we want to measure the strength of social interaction effects as the determinants of individuals educational choices and attainments. The existence of peer-effects is a crucial element for the assessment of various issues linked to the design of an efficient school system: the virtues of an environment where public and publicly funded private schools coexist crucially depend on the size of the peer-effect (Hsieh and Urquiola (2002)); the trade-off between efficiency and equity in allocationg limited resources can be significantly mitigated by explicit consideration of these mechanisms.

To overcome the identification problem outlined above we adopt the instrumental variable approach. We exploit the fact that after the 1980 earthquake that hit some southern Italian regions, the Government relieved from mandatory military service several cohorts of young males in those regions. We speculate that the relief was actually used by these males to increase years of schooling, either by higher university enrollment rates or by higher high school completion rates. Since in Italy females are not subject to military service, the relief had no direct effect on their choices. Therefore, if social interaction played no role in determining individual outcomes we should observe no effect on females' schooling in those cohorts. On the contrary, any observed change can be taken as triggered by social interactions.

Our preliminary results suggest that these effects are actually quite strong. For males the military service relief exogenously increased average education by around 10 weeks; the increase in average females's education in the same group was anything between 7 and 8 weeks.

The paper is organised as follows. Section (2) describes the methodological issues of identification. We then outline our empirical strategy in section (3). Section (4) describes the earthquake and makes the point that it indeed caused an exogenous shift in males' education. Estimates of the social multiplier are given in section (5). Section (6) concludes and outlines further research.

2 A theoretical background: the identification problem.

In this section we briefly discuss the standard identification problem and show how it can be overcome when suitable data ara available. We also show how identification of the parameter can be achieved by looking only at aggregate data. Throughout the section we mainly follow Manski (1993) and adopt his terminology.

We focus on the case where the outcome variable is continuous; in principle, one could look at *decisions*, so that the endogenous variable would be a discrete one (see Brock and Durlauf (2001)).

Let us assume that individuals' outcomes follow a simple general linear rule[1]:

$$y_{ic} = \alpha + \beta X_{ic} + \gamma \bar{X}_c + \lambda Z_c + \theta \bar{y}_c + u_{ic} \tag{1}$$

where X_{ic} are individual exogenous characteristics (such as sex, parental income or education, etc.). X_c collects the averages of these characteristics in the relevant group over which interaction takes place; it captures the fact that individual outcomes may vary with the exogenous characteristics of the group. The parameter γ thus captures the *exogenous* interaction effect. Z_c gathers information about features of the group or the environment in which agents act (e.g. geographical or structural characteristics) which are likely to affect individual behaviour; it thus captures the fact that outcomes may be similar because the environment is; thus λ captures a simple *correlation* effect. y_c is the average outcome in the group and should capture the fact that, all else equal, individual outcomes may be affected by the average performance in the reference group. This *endogenous* interaction effect is captured by θ.

The average equilibrium outcome in group c is then:

$$y_c = \frac{\alpha}{1-\theta} + \frac{\beta+\gamma}{1-\theta}\bar{X}_c + \frac{\lambda}{1-\theta}Z_c$$

which, substituted into (1) yields the following reduced form for individual outcomes:

$$y_{ic} = \frac{\alpha}{1-\theta} + \beta X_{ic} + \frac{\gamma+\beta\theta}{1-\theta}\bar{X}_c + \frac{\lambda}{1-\theta}Z_c + u_{ic} \tag{2}$$

Equation (2) only provides information about the workings of *some* interaction (either *exogenous* or *endogenous*) when the coefficient on $\bar{X}_c$ is different from zero; yet we cannot disentangle one from the other.

Identification of the two parameters becomes possible if the outcome variable of some randomly chosen members of the groups is exogenously altered. In such a situation, θ is identified by the

[1]We assume throughout that average outcome is the relevant variable to look at. Assuming that endogenous interaction occurs through a non linear combination of individual outcomes (say, their variance) would recover directly identification; yet, it would be just out of functional form which, in the absence of any theoretical prior, would be too strong an hypothesis.

exogenous shift of the average outcome in the groups affected by the treatment. Formally, let us augment (1) with a shift variable T_{ic} which equals unity for the N_c^T members of some group c hit by the shock and zero otherwise; δ is by how much the outcome variable has been exogenously shifted. Thus we have:

$$y_{ic} = \alpha + \beta X_{ic} + \gamma \bar{X}_c + \lambda Z_c + \theta \bar{y}_c + \delta T_{ic} + u_{ic}$$

Let n_c be the share of treated individuals in group c, the *strength* of the treatment in group c. In this case the reduced form (2) becomes:

$$y_{ic} = \frac{\alpha}{1-\theta} + \beta X_{ic} + \frac{\gamma + \beta\theta}{1-\theta}\bar{X}_c + \frac{\lambda}{1-\theta}Z_c + \delta T_{ic} + \frac{\delta\theta}{1-\theta}n_c + u_{ic} \tag{3}$$

Equation (3) now allows us to separately identify exogenous and endogenous interactions (γ and θ) provided there is sufficient variability in the share of treated people across groups and T_{ic} does not overlap with other observable individual characteristics included in X_{ic}.

As it will be made clear in the next section, the natural experiment we exploit in the analysis is such that treatment, when given, was addressed only to males. Let us define a dummy T_c only indexed to group, which equals one when treatment was given to males in group c and zero otherwise. Thus rewriting equation (1) separately for males and females:

$$\begin{aligned} y_{ic}^M &= \alpha + \beta X_{ic} + \gamma \bar{X}_c + \lambda Z_c + \theta \bar{y}_c + \delta T_c + u_{ic} \qquad (4) \\ y_{ic}^F &= \alpha + \beta X_{ic} + \gamma \bar{X}_c + \lambda Z_c + \theta \bar{y}_c + u_{ic} \end{aligned}$$

and recalling that group averages are related to sex-group ones by:

$$\begin{aligned} \bar{y}_c &= m_c y_c^M + (1-m_c) y_c^F \qquad (5) \\ \bar{x}_c &= m_c x_c^M + (1-m_c) x_c^F \end{aligned}$$

where we have defined m_c as the share of males in group c, yields the following reduced forms for males' and females' group averages:

$$y_c^M = \frac{\alpha}{1-\theta} + \frac{\lambda}{1-\theta}Z_c + \frac{\gamma + \theta\beta}{1-\theta}\bar{X}_c + \beta X_c^M + \frac{1-\theta(1-m_c)}{1-\theta}\delta T_c \tag{6}$$

$$y_c^F = \frac{\alpha}{1-\theta} + \frac{\lambda}{1-\theta} Z_c + \frac{\gamma + \theta\beta}{1-\theta} \bar{X}_c + \beta X_c^F + \frac{\delta\theta m_c}{1-\theta} T_c$$

The parameter θ could in principle be identified only looking at males. Rewrite the expression for y_c^M as follows:

$$y_c^M = \frac{\alpha}{1-\theta} + \frac{\lambda}{1-\theta} Z_c + \frac{\gamma + \theta\beta}{1-\theta} \bar{X}_c + \beta X_c^M + \delta T_c + \frac{\theta}{1-\theta} \delta(m_c T_c)$$

Identification is achieved by the ratio between coefficients of T_c and $m_c T_c$. Yet, m_c must have sufficient variability across groups in order to avoid multicollinearity among the explanatory variables. Therefore our preferred strategy will imply making the assumption $m_c = 0.5 \quad \forall c$. In this case estimation of equations (6) identifies θ by the ratio of the estimated coefficients on T_c for males and females.

3 An outline of the empirical strategy, data and definitions.

In this section we outline our empirical strategy and introduce some definitions. Let us recall the aim of this paper. We want to assess the existence (and strenght) of endogenous interactions effects in schooling. The previous section has made clear that to identify such a social multiplier in a quite general specification of the decision problem it is crucial that the decision variable of some (identifiable) members of a given group is exogenously shifted; then, any effect on the other members of the same group can only be due to endogenous interaction among individuals.

We think that the earthquake that hit southern Italy in 1980 makes a genuine (yet dramatic) natural experiment suitable to our purposes. Immediately after the earthquake the Government relieved from mandatory military service (MS) all males who were to be drafted in the period 1981-1987 and residing in the earthquake area. The next section describes the event and carefully details the reasons why we can safely use it to our purposes.

Before we can move on to identifying the social multiplier (Section 5) we must fill two gaps. First, identification of a social multiplier requires a definition of the group in which social interaction takes place. Second, we must clearly define what is the variable we think may be affected by such interaction. The rest of this section deals with these two issues.

3.1 Groups.

Throughout most of the paper we mantain the assumption that the group over which interaction takes place is the pool of people born in the same municipality and belonging to the same age cohort. Thus, if the relief from MS caused a shift in males' education, we can identify the social multiplier by looking at what happened to education of females belonging to the same cohort-municipality.

We think this definition of group is a reasonable assumption for several reasons. First, Italian municipalities are generally small: out of 8100 only 6 count a population larger than 500000 thousand. Second, as next section will make clear, the event we exploit to identify these kinds of interaction focuses on individuals aged 14 to 18 thus spending most of their time either in school or anyway with their schoolmates.Both facts undoubtely favour the workings of the kind of interaction we are after.

3.2 Measures of schooling.

So far we have been loose in defining what we mean by *schooling* or *education*. The ideal quantitative measure would be years of school attendance. Unfortunately our main data source, the 1991 population census, only reports information on completed schooling at the survey date. In 1991 the cohorts relieved from military service were aged 25 to 29, with most of them having finalised their education. Yet, actual years of education are likely to be underreported: on the one hand we miss all schooling taken by people still in school at the survey date and on the other all schooling taken by drop-outs on top of their declared completed education. In order to mitigate this measurement problem, we assign some years of education on top of the declared competed level to all those individuals that in the 1991 census reported being still students. We impute to these individuals $(AGE - 6)$ years of schooling, on the grounds that primary school starts at the age of 6.

Another way of looking at the problem is studying the *decision* of attending university. We would then like to have a measure of enrollment of people belonging to any given cohort-municipality. Again, it is not available (we do not know whether a given individual ever attended university unless she completed it). This measure would be more reliable at capturing *imitation* effects or the simple transmission of information since it only considers the *decision* of attending university which is likely to have been taken in the municipality of birth. The quantitative measure could instead also capture, say, university fixed effects (better vs. worse universities attended, etc.), all effects which although interesting, we could not identify unless we were sure that all individuals of a given municipality-cohort attended the same university (or department, etc.). Yet, the situation is not so bad since, although there are serious reasons why we could not safely assume that all people of a given group attended the same university, it is also true that that there are serious reasons to assume that the pattern of mobility of teenagers to attend university are not significantly different between control and treated municipalities; if they were then our estimates could mistake a, say, university effect, for a social interaction effect.

Another problem is related to the requirement that people were actually residing in the earthquake regions to benefit from the relief from MS. This information is not available in the census, where only current residence and place of birth are collected. We then choose to proxy the place of residence at the time of the earthquake with the place of birth and assume that all people in the relevant cohorts and born in those municipalities comply with the requirement. This assumption would be a problem if people aged 14-18 in 1980 and born in the earthquake municipalities have a higher probability of being resident somewhere else in 1980, thus not being allowed to exploit the relief. Then, if, say, their education turns out to be higher than in the comparison group, this could not be taken as evidence that the relief increased schooling. The next section deals also with this issue.

4 A natural experiment? The 1980 earthquake.

On the evening of November 23rd 1980 southern Italy was hit by an exceptional earthquake. Around 650 municipalities and more than 5 milion persons were affected by the seismic event. In figure (**??**) we show the area and, the darker region, the epicentre of the seismic event. Earthquakes continued to hit until late February 1981.

The Government took immediate action to sustain the population hit by the seismic episodes. Several intervention plans were rapidly passed. Among those, the relief from mandatory military service for all males born in the period 1962-66 with legal residence in the municipalities hit by the earthquake.

While all kinds of financial help were related to measures of damage established by a commission of experts appointed by the Ministry of Internal Affairs, the relief was granted independently of the damage borne by the individual, his family or his municipality.

4.1 Did males' education change because of the MMSR?

A relief from mandatory military service does not, by itself, increase school attendance or completion rates. Individuals may just behave, as concerns schooling, as they would in the absence of the relief. For example, someone who drops out from upper secondary school before completion because of the draft and then starts working may just as well drop out at the same age and immediately start working.

We must then be sure that the MMSR actually determined a change in the schooling of males who were granted it. If this turns out to be the case, then an exogenous shock (the earthquake) caused a shift in the endogenous variable and we can use it as an instrument to identify the social multiplier.

We have seen that not all municipalities in the earthquake region were damaged to the same extent: those closest to the epicentre borne the largest costs while those along the border of Campania and Basilicata were generally not significantly affected. The cohorts of 1962-66 of males living in these municipalities were nonetheless given the opportunity of skipping the military service.

To assess the extent to which the relief altered males' schooling in the municipalities involved, we need a benchmark group. In selecting treated and control municipalities we attempted to replicate what would have been the outcome of a random assignment of the treatment (the quake). Therefore we selected a portion of the Italian territory that included both type of municipalities, that beside being selected and not, did share the same characteristics. To implement this selection rule we choose all the municipalities on the boundaries of the areas that benefited of the relief of the Mandatory Military Service, and for the control group all the municipalities that were neighbourhood of at least one member of the treated group. We end up with 117 municipalities belonging to 12 provinces (Benevento, Caserta, Foggia, Matera, Potenza, Bari, Campobasso, Cosenza, Frosinone, Isernia, Latina, Taranto) of 5 regions (Lazio,Molise, Puglia Campania, Basilicata, Calabria) with a total population at the end of 1979 of about 900.000 people. Figure (**??**) shows the treated and control municipalities selected, the former in darker grey. The detailed list of the selected towns are reported in the appendix.

This choice has several advantages. First since treated municipalities are rather peripheral with respect to the centre of the earthquake they suffered little damages, very much like the non treated municipalities. Table (1) presents a synthetic indicator of the damages suffered by the treated municipalities. It combines 6 indicators of damage that take into account specific damages (number of people who died or were injured in the quake, number of people who lost their house, the index of the damage computed by the Ministry of Internal Affairs, number of houses destroyed, number of houses damaged, number of temporary houses needed). The synthetic indicator summarise these six indexes into a scale that ranges from 6 to 30: a municipality would score 6 if there were no deaths or injured people, the homeless were less than 2 per cent of the population, the index of the Ministry of Internal Affairs was less than 5 per cent, the destroyed houses were less than 0,5 per cent those damaged less than 1 per cent, and temporary shelter was needed for less than 5 per cent of the population. At the other extreme a municipality would score 30 if at least 10 per cent of the population died or was injured, at least 20 per cent were homeless, the index of the Ministry of Internal Affairs was at least 70 per cent, more than 30 per cent of the houses were destroyed or heavily damaged and temporary shelter was needed for more than 40 per cent of the population. The synthetic indicator has been grouped into 6 classes as indicated in the Table. For each class we report the number of municipalities involved and few basic statistics (mean, max, min, and standard error) of their synthetic index. Only 49 of our 57 treated municipalities are reported in the table because several of them were considered outside of the area involved by the quakes (yet they were included in the group that enjoyed the lift of the MMS.). Only two out of 49 municipalities suffered serious damages as they were included in the third group and had a average score of 18. Another group of 16 municipalities are in group 3 with an average score of 12,7. The rest of the municipalities are included in the last two group, 14 of them with a mean score of 8,3 and 17 of them with a mean score of 6.

The second advantage of our choice is that by including all municipalities on the boundaries that

Table 1: Assessment of damages.

Class of damage (Range 6-30)	Number	Mean score	max score	min score	se
1) very highly damaged: municipalities in the Cratere					
	0	0	0	0	0
2) highly damaged: municipalities outside the Cratere with score$\geq$ 20					
	0	0	0	0	0
3) Score 16-20					
	2	18,0	20	16	2,0
4) Score 11-15					
	16	12,7	15	11	1,2
5) Score 7-10					
	14	8,3	10	7	1,0
6) Score 6					
	17	6	6	6	0

stretch in the direction north-south we are actually selecting into our sample a wide variety of people. Thus our results should be held to be quite general and not depending on the choice of a very narrowly defined group. Our sample allows us to claim that our result would apply to (at least) a large portion of the Italian population.

The third advantage refers to comparability of the treated and control group. Since each municipality in the control group neighbours with at least a treated municipality, the two group should roughly share the same characteristics. In Table (2, 3 and 4), we look at three different groups of characteristics and check how they differ across treated and controls municipalities. The first group, examined in table (2), refers to structural characteristics. We report the distribution of the municipalities across provinces, their size and some physical attributes. Although treated municipalities appears on average smaller in terms of squared kilometers (382) than control (459) there is a great deal of heterogeneity in each group so that their difference is not statistically different than zero. Same message is delivered when we look at the population as measure of the size. Treated municipalities are on average smaller than the control (4960 people versus 10245) however the dispersion with group is such that the difference of the mean size is hardly different than zero. More evidence of systematic diversity between treated and control is delivered by the population density that suggests that people in treated municipalities are more sparse (13 people per square kilometer) than the other group (18 people per square kilometer). The difference seems to be statistically significant. Notice that this result make more difficult to find social interaction in the treated group. Although a larger share of the treated municipalities is located in mountains the average altitude does not differ between treated and control.

Table (3) reports population characteristics. Overall the larger differences seem to be limited to births and death rate. Treated and control municipalities are more alike along gender composition, mobility patterns and age composition of the population. In 1979, 283.000 people were living in the treated municipalities and 615.000 in the control. People born alive in the two groups of municipalities are almost the same (1,5 and 1,6 per cent of the population) but the difference is still statistically different from zero because of the small within-group variance. The fertility rate is smaller in the treated group (2,9 per cent of female population) than in the control (3,1 per cent) and the distance is statically large. In 1979, 9 people died over 1000 in the treated group, only marginally, and yet significantly, more than the control group. Males are slightly less than females in both groups; mobility patters seems to be alike across treated and control group.

Table (4) deals with school characteristics; the main message delivered by the table is that up to 8 grade, treated municipalities, in comparison with control, have less student as a ratio of the population and a higher endowment of both schools and rooms. Therefore have smaller school and class size. We need to take into account those differences as long as these differences may influence the number of years of schooling achieved by the population. However good news come from high schools where the differences disappear altogether; treated and controls are all alike along the all dimension; they have an equal number of students in the population, same amount of school, same density and the same ratio between high-school students versus student enrolled in grade lower than 9th.

We conclude that the municipalities closest to the border of the earthquake area, yet outside it (Lazio, Puglia, Calabria) make a suitable control group for the municipalities lying along the inner border of the earthquake area. The facts shown above make us sure that these two groups of municipalities

Table 2: Structural characteristics of municipalities in the sample.

	Tretaed		Control		All		Treated vs Control	
	Mean or count[a]	share or sd	Mean or count[a]	share or sd	Mean or count[a]	share or sd	coeff	tstat
Total	57		60		117			
Benevento	12	0.211			12	0.103		
Caserta	15	0.263			15	0.128		
Foggia	5	0.088	12	0.2	17	0.145		
Matera	8	0.140			8	0.068		
Potenza	17	0.298			17	0.145		
Bari			7	0.117	7	0.060		
Campobasso			9	0.150	9	0.077		
Cosenza			15	0.250	15	0.128		
Frosinone			5	0.083	5	0.043		
Isernia			7	0.117	7	0.060		
Latina			3	0.050	3	0.026		
Taranto			2	0.033	2	0.017		
Size								
Land (km2)	382.7	386	458.9	583	421.75	496	-76.2	-0.83
People (1980)	4961	7554	10245	22475	7671	17073	-5284	-1.69
Pop. dens.	12.8	8.5	17.8	16.52	15.4	13.42	-5	-2.04
Location								
Internal Mount.	29	0.51	17	0.283	46	0.393		
Costal Mount.	2	0.04	1	0.017	3	0.026		
Internal Hills	23	0.40	30	0.500	53	0.453		
Costal Hills	1	0.02	6	0.100	7	0.060		
Flat Land	2	0.04	6	0.100	8	0.068		
Altitude[b]	496.1	225	456.1	265.9	475.6	246.6	40.09	0.88
Costal Mun.								
-No	53	0.930	53	0.883	106	0.906		
-Yes	4	0.070	4	0.067	8	0.068		
Less than 5 km from the coast	0	0	3	0.050	3	0.026		

[a]) Population weighted
[b]) meters above sea level

Table 3: Population characteristics at 31 Dec. 1979.

	Tretaed		Control		All		Tretaed vs Control	
	Mean or count[a]	sd	Mean or count[a]	sd	Mean or count[a]	sd	coeff	tstat
Total population	282780		614742		897522			
Born alive/Population	0.015	0.003	0.016	0.003	0.016	0.003	-0.002	-2.91
Born alive/Female	0.029	0.006	0.032	0.006	0.031	0.006	-0.004	-3.16
Deaths/Population	0.009	0.002	0.008	0.002	0.008	0.002	0.001	2.36
Males	0.494	0.054	0.494	0.007	0.494	0.031	0.000	-0.01
Inflow	0.017	0.006	0.020	0.011	0.019	0.010	-0.003	-1.35
-from abroad	0.002	0.002	0.002	0.002	0.002	0.002	0.000	-0.08
Outflow	0.023	0.009	0.021	0.007	0.022	0.008	0.002	1.43
-abroad	0.002	0.002	0.001	0.002	0.001	0.002	0.000	0.11
Age composition								
(6-14/all population)	0.140	0.047	0.151	0.019	0.147	0.031	-0.010	-1.66

[a]) Population weighted

make a homogeneous population and that the MMSR granted after the earthquake to some of them can actually be seen as a treatment to these municipalities. For the sake of brevity we will often refer, when no confusion can be generated, to the municipalities in the earthquake area as treated and to the ones outside as control.

We now turn to the crucial question. Did males use the relief to go more to school? Figure (2) answers this question and adds arguments to the reliability of the comparison between the two groups of municipalities. On the vertical axis we report average years of males' schooling in the two groups of municipalities; on the horizontal axis there is age in 1991; schooling of the control group is the full line. It is striking that the average levels of schooling are nearly identical for all the cohorts aged 30-50, while they move apart for the cohorts aged 25-28, those directly *treated*.

Taken together the evidence presented in this section makes us sufficiently confident that the difference between males' schooling for the cohorts 1962-66 in the two groups of municipalities is largely exogenous. We can thus use the earthquake and subsequent military relief as a valid instrument to identify the workings of social interactions. This is done in the next section.

Table 4: School characteristics, 1979.

	Tretaed		Control		All		Tretaed vs Control	
	Mean or count[a]	share or sd	Mean or count[a]	share or sd	Mean or count[a]	share or sd	coeff	tstat
Primary schools								
Student/pop.	0.084	0.024	0.094	0.014	0.091	0.019	-0.011	-2.93
Schools/pop.	0.001	0.001	0.001	0.001	0.001	0.001	0.000	3.03
Rooms/pop.	0.005	0.002	0.004	0.001	0.004	0.002	0.001	4.41
Student/school	198.4	194.8	350.5	225.4	302.6	226.7	-152.1	-3.54
Student/room	16.5	4.714	25.4	7.890	22.6	8.142	-8.8	-6.31
Junior High school								
Student/pop.	0.056	0.024	0.056	0.007	0.056	0.015	0.000	0.13
Schools/pop.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.59
Rooms/pop.	0.003	0.001	0.003	0.000	0.003	0.001	0.000	2.32
Student/school	306.3	152.8	444.1	160.2	400.7	169.8	-137.8	-4.25
Student/room	19.5	3.197	21.5	2.748	20.9	3.037	-2.0	-3.43
High Schools								
Student/pop.	0.047	0.052	0.049	0.042	0.049	0.046	-0.002	-0.27
Schools/pop.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.40
Rooms/pop.	0.003	0.003	0.003	0.023	0.003	0.002	0.000	-0.03
Student/school	252.3	156.4	326.6	187.4	306.6	180.4	-74.3	-1.10
Student/room	17.9	2.402	18.7	3.336	18.5	3.091	-0.7	-0.61
HS/(JHS+PRI)[b]	0.321	0.348	0.321	0.283	0.321	0.304	0.000	0.01

[a]) Population weighted

[b]) High School/(Junior HS + Primary) students.

Table 5: Cohorts 1962-1966.

	MALES		FEMALES		ENDOGENOUS INTERACTION	
	b_M	S.E.	b_F	S.E.	$\hat{\theta}$	S.E.
25-28	0.3028	0.1015	0.4623	0.1094	1.21	0.1127

5 The social multiplier.

We are now ready to undertake the main task of the paper and provide estimates for the endogenous interaction coefficient. Our dependent variables will be average males' and females' years of schooling in each of the 117 municipalities considered. We will estimate several specifications of the two reduced forms (6) separately and then determine the implied endogenous interaction coefficient θ. All estimates are weighted by the number of observations in the municipality-age cell.

We start out providing estimates for the two reduced forms obtained only on the sample of cohorts 1962-66, that is those relieved from MS in the municipalities hit by the earthquake.

Since the shock is truly exogenous (the earthquake) we can safely estimate this specifications not worrying about individual as well as group and municipality controls $(X_{ic}, \bar{X}_c, Z_c)$. Therefore we estimate:

$$\begin{aligned} y_c^M &= a_M + b_M E_c \qquad (7) \\ y_c^F &= a_F + b_F E_c \end{aligned}$$

where E_c equals one for the municipalities c where males were relieved from MS. Recall that we identify θ by means of the ratio between males' and females' reduced form coefficients on E_c, under the working assumption that the share of males equals one half in all municipalities. The assumption is not crucial since it is not too far away from true data. Moreover our main source of identification is the endogenous effect on females and not the variability of treatment across municipalities. Results are shown in table (5). The estimated b_M for males says that on average males' schooling went up by roughly 4 months becasue of the military service relief. The coefficient is precisely estimated and strongly significant. The effect of males' relief from MS on females (b_F) is actually stronger, roughly 6 months of additional schooling. These estimates for the reduced form coefficients imply a strong effect of endogenous interaction. Actually this effect is even too strong: the social multiplier being greater than unity, an equilibrium in the simple model shown in section (2) would not exist. The problem with these estimates is that we are assigning all differences in females' education across municipalities to the workings of endogenous interaction. We want to be more strict and allow for possibly different average education levels across the two groups of municipalities. We do this by adopting a discontinuity design approach: we include in our sample also slightly older age groups, that is those who were not relieved from military service even if they belonged to the municipalities hit by the earthquake. In the first specification we estimate, we introduce a further control T_c which equals one for all age groups belonging to municipalities hit by the earthquake, even if they were not relieved from MS; this allows us to control also for a possible different intercept in average education across the two groups of municipalities. The reduced form differential effects due to military relief are again captured by a dummy E_c which now equals unity only for the age-municipality groups relieved from MS. The second more flexible specification we estimated allows for a municipality-specific intercept, G_c. Again in both specifications identification of θ obtains from the ratio between b_M and b_M. As a very preliminary robustness check we run both specifications on different samples, including more and more cohorts, as it will be clear below. In both specifications possible secular increase in education is controlled for by including also age among the regressors. Results for the relevant coefficients are reported in tables (6) and (7); the first column of both tables reports the age groups included in the samples; the last two columns report the implied θ and its standard error, computed by means of a Taylor expansion around the estimated values for b_M and b_M (delta method). The reduced form coefficients $\{b_M, b_F\}$ turn out to be very stable across samples for both specifications, more so for males, and gain significance as sample size increases. Moreover, they are lower than those reported in table (5): the reduction is much larger for women, signalling that indeed the age control plays a role in explaining part of the difference between education levels of females in the cohorts 1962-66 in the two groups of municipalities. Although both coefficients turn out to be lower when municipality-specific dummies are introduced, the implied θs remain almost unchanged and statistically significant. These results

Table 6: Model: $y_c = a + bE_c + dT_c + hAGE_c$.

	MALES		FEMALES		ENDOGENOUS INTERACTION	
	b_M	S.E.	b_F	S.E.	$\hat{\theta}$	S.E.
28-29	0.180	0.267	-0.02	0.293	-0.25	4.14
27-30	0.166	0.183	0.090	0.198	0.703	1.12
26-31	0.198	0.149	0.164	0.160	0.906	0.61
25-32	0.204	0.130	0.179	0.138	0.935	0.50
25-33	0.212	0.123	0.197	0.131	0.963	0.44
25-34	0.206	0.117	0.181	0.126	0.935	0.45
25-35	0.201	0.112	0.196	0.120	0.987	0.41
25-36	0.209	0.109	0.189	0.116	0.950	0.40

Table 7: Model: $y_c = bE_c + dG_c + hAGE_c$.

	MALES		FEMALES		ENDOGENOUS INTERACTION	
	b_M	S.E.	b_F	S.E.	$\hat{\theta}$	S.E.
28-29	0.191	0.117	- 0.028	0.139	0.344	2.014
27-30	0.108	0.083	0.060	0.088	0.714	0.760
26-31	0.131	0.068	0.124	0.071	0.973	0.386
25-32	0.131	0.061	0.114	0.065	0.931	0.366
25-33	0.148	0.060	0.136	0.062	0.958	0.304
25-34	0.139	0.057	0.118	0.059	0.918	0.321
25-35	0.134	0.056	0.134	0.057	1.000	0.298
25-36	0.146	0.055	0.125	0.056	0.923	0.291

Table 8: Controlling for pre-trends.

	MALES		FEMALES		ENDOGENOUS INTERACTION	
	b_M	S.E.	b_F	S.E.	$\hat{\theta}$	S.E.
25-50	0.217	0.078	0.061	0.080	0.439	0.217
25-50	0.730	0.082	0.400	0.088	0.707	0.013

are now consistent with existence of an equilibrium in the simple model introduced in section (2).

There are two main drawbacks with these results. First, we are controlling for a *common* time trend for all municipalities. Inspection of figures (2) and (3) reveals that this is not a bad approximation for males, less so for females. Second, the time trend is estimated only on a small set of age groups, at most 11 data points for the sample 25-36. Therefore it is likely to be driven in a significant way by what happens in the age groups where endogenous interaction can be identified, the more so the stronger the effect of this kind of interaction. We tackle this issue by following a standard approach in the literature (see, among others, Ashenfelter and Card (1985)): we build a counterfactual by estimating a time pattern for education on the age groups not touched by the treatment (29-50); we adopt a very flexible specification, allowing for a different quadratic trend and intercept for the two groups of municipalities and estimating again separately for men and women. We then extrapolate this structure to the age groups left out, that is the 25-28 bracket for both sexes and all groups of municipalities. We then regress the differences between actual and fitted education on the dummy E_c as defined above. Results are reported in table (8). The first row reports estimates where the fitted values are obtained using all coefficients; in the second one we have fitted by dropping 5% non statistically significant coefficients. Results show very different estimates for θ, although both are 5% significant.

6 Conclusions and open questions.

In this paper we tackled the issue of identifying the workings of endogenous social interactions in schooling, that is the dependence of individual outcomes on aggeregate ones in a way not mediated by markets. Rather than relying on ad hoc assumptions on the structure of the problem to be studied, we build a natural experiment from the 1980 earthquake that severely hit some southern Italian regions and the subsequent relief from mandatory military service granted to all males aged 14 to 18 living in the regions hit by the seism. We show that males exploited this relief by increasing their average years of schooling by roughly ten weeks. Italian females are not subject to mandatory military draft therefore the militart relief granted after the earthquake had no direct effect on their choices. Nonetheless we find that their average years of schooling went up by 7 to 8 weeks. This can only be due to the workings of some kind of endogenous social interaction. These results, although preliminary, are robust to the choice of the cohorts chosen as control group. Yet, further robustness check are due. We plan to perform the analysis on subsets of the municipalities belonging to our sample, to include in our regressions family controls, to experiment with a different measure of schooling by looking at university enrollment rates. We think that these controls may be relevant because:

1. although we have shown that treated and control municipalities are on average very much alike, the variances of characteristics and average outcomes within each group are large. Therefore, we need to shrink both groups so as to make the comparison tighter;

2. we have run our regressions with no individual controls. This is legitimate since the shock is fully orthogonal to any individual or municipality characteristics, delivering a consistent and unbiased estimate for the reduced forms parameters. However, the estimated standard errors are large implying, in some cases, non significant estimates. A way to fix this problem is to perform the analysis at the individual level using a full set of individual controls;

3. average years of schooling as measured in the paper miss all the information about college drop-outs. Yet, if social interaction affects the *decision* whether to enroll at college then looking at average completed years of schooling underestimates the strength of endogenous interactions.

References

[1] Ashenfelter O. and D. Card (1985) - "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs", Review of Economics and Statistics, 67(4).

[2] Brock W. and S. Durlauf (2000) - "Interaction-based Models", NBER Technical Working Paper n. 258.

[3] Brock W. and S. Durlauf (2001) - "Discrete Choice with Social Interactions", Review of Economic Studies, 68(2).

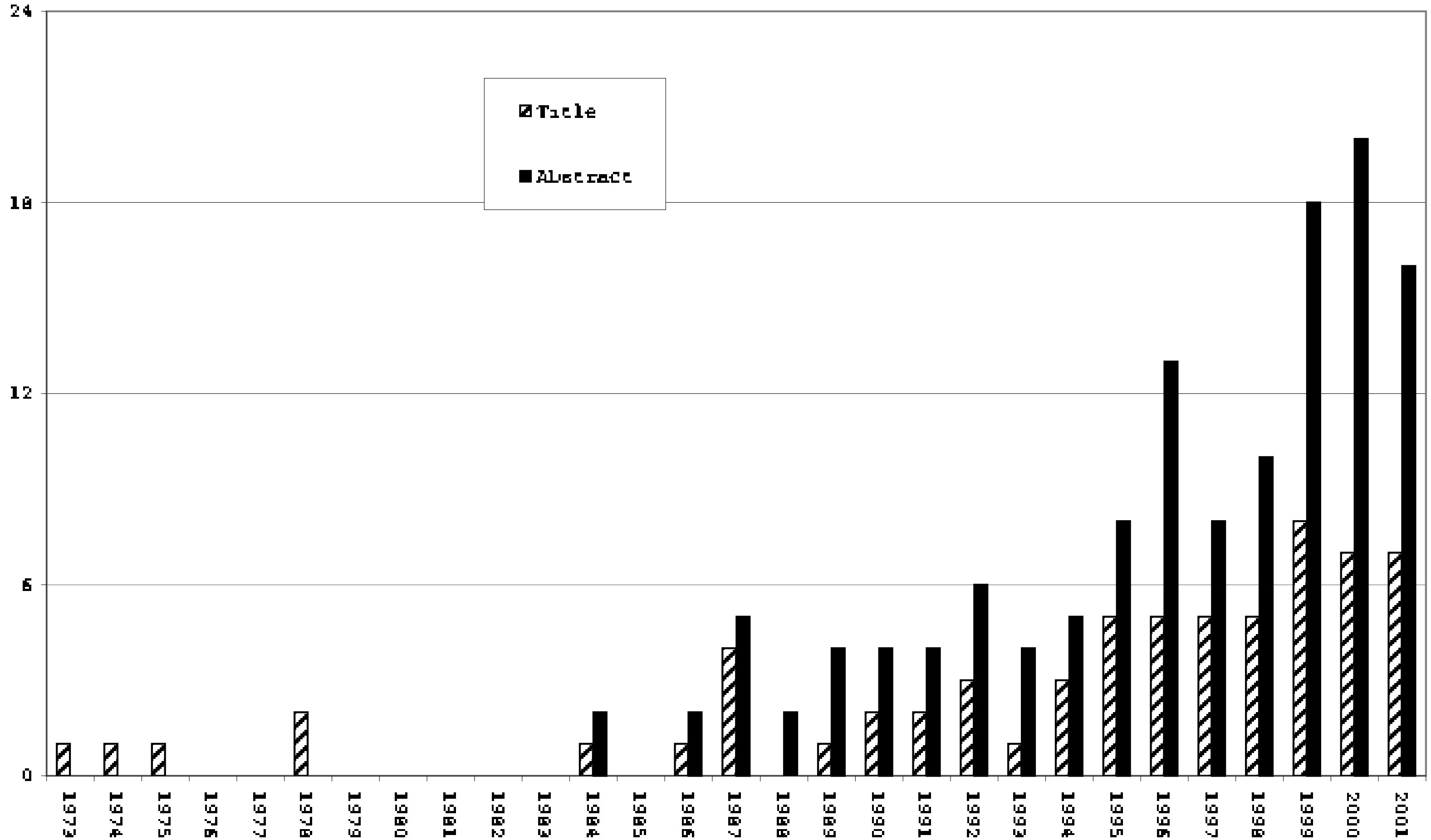
[4] Hsieh C. and M. Urquiola (2002) - "When Schools Compete, How do They Compete?", mimeo.

[5] Manski C. (1993) - "Identification of Edogenous Social Effects: The Reflection Problem", Review of Economic Studies, 60(3).

[6] Ministero del Bilancio e della Programmazione Economica (1981) - "Rapporto sul Terremoto".

[7] Moffitt R. (2000) - "Policy Interventions, Low-Level Equilibria and Social Interactions", mimeo.

Figure 1: EconLit hits.



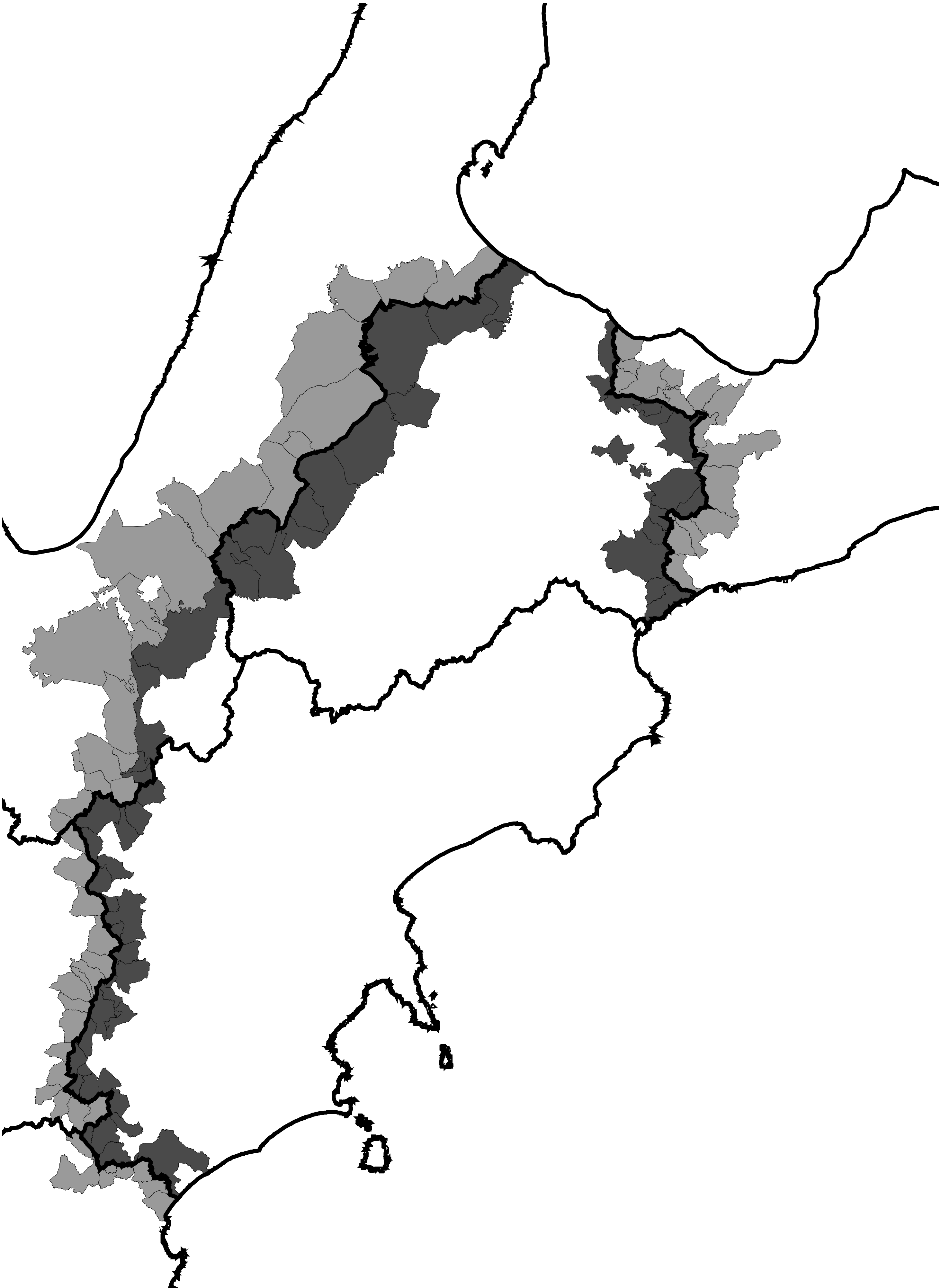


Figure 2: Average years of schooling by age in 1991: males

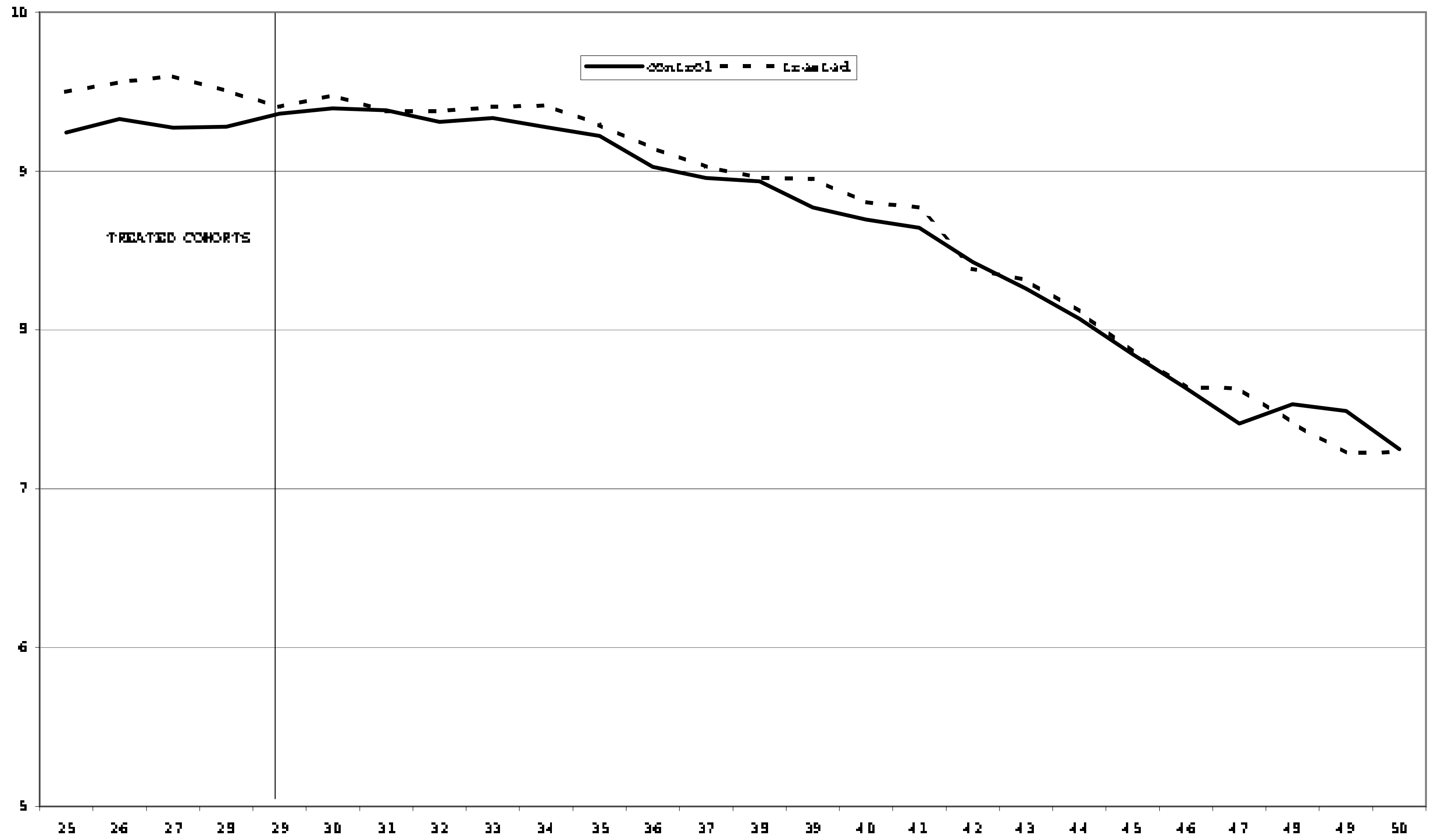


Figure 3: Average years of schooling by age in 1991: females