NBER WORKING PAPER SERIES

UNDERSTANDING SOCIAL INTERACTIONS:
EVIDENCE FROM THE CLASSROOM

Giacomo De Giorgi
Michele Pellizzari

Working Paper 19202
http://www.nber.org/papers/w19202

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2013

We thank Richard Blundell, Pat Bayer, David Card, Joan de Marti, Matt Harding, Caroline Hoxby, Matt Jackson, Pat Klein, Thomas Lemieux, Enrico Moretti, Eleonora Patacchini, Jesse Rothstein and seminar participants at UC-Berkeley, NBER SI 2010, UCLA, UC-Davis, Universitá di Bologna, Stockholm University, Uppsala University. We are also grateful to Bocconi University for allowing access to their archives. In particular Giacomo Carrai, Mariele Chirulli, Mariapia Chisari, Alessandro Ciarlo, Enrica Greggio, Gabriella Maggioni, Erika Palazzo, Giovanni Pavese, Cherubino Profeta, Alessandra Startari and Mariangela Vago have all been a constant source of precious help and information. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Understanding Social Interactions: Evidence from the Classroom
Giacomo De Giorgi and Michele Pellizzari
NBER Working Paper No. 19202
July 2013
JEL No. I21,J0

ABSTRACT

Little is known about the economic mechanisms leading to the high level of clustering in behavior commonly observed in the data. We present a model where agents can interact according to three distinct mechanisms, and we derive testable implications which allow us to distinguish between the proposed mechanisms. In our application we study students' performance and we find that a mutual insurance mechanism is consistent with the data. Such a result bears important policy implications for all those situations in which social interactions are important, from teamwork to class formation in education and co-authorship in academic research.

Giacomo De Giorgi
Departament d'Economia i d'Historia Economica
Edifici B, Office B3-1122
Universitat Autònoma de Barcelona
08193 Bellaterra, (Barcelona)
Spain
and NBER
giacomo.degiorgi@gmail.com

Michele Pellizzari
University of Geneva
Uni Mail, Boulevard du Pont-d'Arve 40,
CH-1211 Geneva 4, Switzerland
and IZA
michele.pellizzari@unige.ch

1 Introduction

There is a firm belief that social interactions are important determinants of behavior in a variety of contexts (Matthew O. Jackson 2008, Matthew O. Jackson 2009), yet very little is known about the mechanics of such interactions. We propose and test models of social interactions where risk-averse agents engage in production and can either (i) act in isolation, (ii) exploit possible complementarities (joint production), or (iii) mutually insure against idiosyncratic shocks to time or productivity.

These models lend testable predictions in the first and second moments of the distribution of outcomes that allow us to distinguish them in empirical applications. In particular, while all three mechanisms generate outcomes that are correlated within groups, they bear different implications for the average performance of group members.

We consider the model in which agents act in isolation, but may care about each others' performance, as a benchmark and we compare the implications of joint production and insurance against it. We show that under joint production agents exploit complementarities to achieve a higher level of output compared to the benchmark. In the mutual (full) insurance scenario, any idiosyncratic shock to productivity (or endowments) across agents is eliminated and the within-group comparisons that induced higher effort in the benchmark case disappear. As a consequence average output is lower in mutual insurance than in the benchmark case.

Despite the fact that our three mechanisms generate different predictions for average performance, they all lead to clustering of outcomes within groups, which is what is commonly referred to as peer effects or social interactions. Under decentralization (i.e. our benchmark model) such a result rests on the assumption that agents evaluate their performance relative to their peers, as in a tournament or status seeking model (Sherwin Rosen 1986, George A. Akerlof 1997). Namely, the utility one enjoys out of a good performance declines if everyone in the group does well. A recent paper by David Card, Alexandre Mas, Enrico Moretti & Emmanuel Saez (2012) provides convincing evidence that relative utility is important to workers as they learn about their co-workers' compensations. When agents produce jointly, the effort levels of all members of the group enter each other's production functions, thus inducing correlation in outcomes. Finally, in the mutual insurance setting the heterogeneity due to the idiosyncratic shocks is eliminated, thus mechanically reducing the dispersion of outcomes within groups.

In our model agents choose their optimal level of effort conditional on the idiosyncratic shock, while the choice of the interaction mechanism is made beforehand under uncertainty according to a standard von Neumann-Morgenstern expected utility function. We further assume that there are costs associated to active social interactions, such as those required to produce jointly or to engage in mutual insurance, and that such costs decrease as agents get to know each other. Consistently with this framework, in our empirical application we examine how the mean and the dispersion of outcomes evolve with the frequency of students' meetings in the classroom.

We test the implications of our model using data on undergraduate students at Bocconi University, who are randomly assigned to teaching classes. Moreover, the random allocation is repeated at the beginning of each academic year, so that the data exhibit a large degree of exogenous variation (both cross-sectionally and over time) in the number of hours any two students spend together in the same classroom. We exploit such exogenous variation to test the implications of the interaction mechanisms considered in our theoretical discussion on the mean and the variance of academic performance, both cross-sectionally and over time.

Our results indicate that mutual insurance is the economic mechanism that prevails in the setting of our application. Specifically, we find that pairs of students who are more often allocated to the same classes are characterized by less dispersed outcomes and lower average performance (both cross-sectionally and over time). Of the three mechanisms considered, insurance is the only one capable of producing both of these results. As an additional test of insurance, we also find that the likelihood that the grade vectors of any two students cross decreases with the number of hours they have been randomly allocated to the same classes. Such a result is consistent with the mutual insurance mechanism, which preserves the ranking of agents' outcomes across states and time.

In terms of magnitude, our results suggest that the mean grade of students who attend all their compulsory courses together would be 12% of a standard deviation (0.2 grade points) lower that that of students who attend each course with an entirely new group of peers.

To our knowledge, we are amongst the first to explicitly investigate the economic nature of social interactions. The literature has been dominated by the search for suitable identification strategies to solve the many econometric hurdles of peer effects models (Charles F. Manski 1993, William A. Brock & Steven N. Durlauf 2001, Robert Moffitt 2001) and has, so far, devoted very little attention to understanding the different mechanisms that may generate such effects.[1] One notable exception is Jane Cooley (2009), who looks at a series of theoretical models to derive the empirical functional forms of spill-overs across classmates. A different approach is taken by Alberto Bisin, Andrea Moro & Giorgio Topa (2011), who discuss the identification of models of social interactions in the presence of multiple equilibria, although their analysis is carried out in a setting where the economic mechanism is not directly investigated.

We test the implications of our models in the context of higher education but the set-up and the topics discussed in this paper are quite general and can be applied to many other situations. For example, in a production team the pressure exerted by the social group, either through sanctions or by simple relative utility, may alleviate the free-riding problem (Eugene Kandel & Edward P. Lazear 1992, Armen Alchian & Harold Desmetz 1972, Bengt Holmstrom 1982). Peer pressure also explains the results of Alexandre Mas & Enrico Moretti (2009), where supermarket checkers' performance improves when they are paired with high performance colleagues. Similarly, peer pressure seems to provide a consistent explanation also for the findings in Armin Falk & Andrea Ichino (2006).[2]

A key difference between our theoretical approach and the typical model of team production is the verifiability of individual output by the principal. Consistently with the empirical application, in our model individual performance is verifiable, there are no (explicit) principals nor informational asymmetries, however agents face an uncertain environment characterized by random shocks to their time endowments or to their productivities. We believe that many production problems resemble the one

[1] A recent wave of studies has approached the identification problem with the use of network structures (Yann Bramoullé, Habiba Djebbari & Bernard Fortin 2009, Coralio Ballester, Antoni Calvó-Armengol & Yves Zenou 2006, Giacomo De Giorgi, Michele Pellizzari & Silvia Redaelli 2010).

[2] Oriana Bandiera, Iwan Barankay & Imran Rasul (2009a) show how social connections affect productivity in teams; while Patrick Bayer, Randi Pintoff & David Pozen (2009) provide evidence on the underlying nature of the interaction in criminal behavior, suggesting a mechanism based on information or learning from others.

studied in this paper, from manufacturing to construction, to most white collar jobs or coauthoring in academic research, where tenure decisions are made at the individual level but the research activity is often carried out within groups of co-authors.

Without knowledge of the mechanics of social interactions it is often impossible to foresee the effects of policy interventions. A notable example is Scott Carrell, Bruce I. Sacerdote & James E. West (2012), whose experiment on optimal sorting of students based on reduced form parameters leads to unexpected and unintended results. In this view, the policy implications of our paper are far-reaching. In our application, for example, if the aim of the policy maker is to maximize average performance, one would design a mechanism of class allocation that prevents students from meeting too frequently and/or introduces incentives to favor joint production and limits the possibility of mutual insurance.

Whenever production is organised in teams, knowing that agents may engage in mutual insurance is important for the principal, who may design a process of team formation and rotation that maximises the benefits of cooperation without encouraging insurance behaviours (Kandel & Lazear 1992, Bandiera, Barankay & Rasul 2009a, Oriana Bandiera, Iwan Barankay & Imran Rasul 2012).

The interpretation of our findings in the light of a well-defined model of mutual insurance allows us to draw implications for a number of other settings, thus reinforcing the external validity of our analysis. However, viewed from the strict policy perspective of education institutions that are similar to Bocconi university, our results bear important implications regardless of the exact nature of the interaction mechanism. Knowing that average performance may decrease when interactions are particularly frequent is sufficient to suggest reallocating students to classes in a way that can prevent such a decline. At the same time learning what's the mechanisms can inform the school on the policy action to increase mean and reduce variances both across students and overtime. The college can still decide to keep the current allocation while also introducing productivity prizes and bonuses, that could ensure a low variances of grades as well as revert the fall in the grades due to insurance. Importantly, if the school were to provide standardized teaching material that would per se reduce the need for peers' insurance, but the fall in grades would still persist even if the number of meetings were to be dropped.

We admittedly propose and test a limited set of models and some of our results can certainly be captured by alternative mechanisms.[3] Our analysis more modestly shows that the patterns of observed social interactions can be explained by a small set of simple economic mechanisms, that are relatively standard and well-established in the literature.

In this sense, we depart significantly from most of the literature on peer effects and social interactions, that focuses on the identification and estimation of the linear-in-means model (Manski 1993). As acknowledged by most authors, that model can be seen as the econometric counterpart of several theoretical mechanisms of interactions and one reading of our contribution is precisely the attempt to uncover such mechanism, within a small set of standard economic models.

The structure of the paper is as follows: Section 2 presents our three simple models of social interactions (decentralization, joint production and mutual insurance); in Section 3 we provide a description of the data used in Section 4 to carry out the empirical exercise and take the predictions of the model to the test. Section 5 briefly discusses alternative mechanisms and, finally, Section 6 concludes.

[3] See Section 5 for a discussion of alternative models.

2 Model

In this section we present a simple model of social interactions. For expositional purposes and coherence with the empirical application, we will frame the model in the education environment and characterize the two agents as students and the production process as learning. Otherwise, the model can be applied to other settings and it is meant to parsimoniously include the key features of several interaction mechanisms and to produce testable implications that can be taken to the data.

Two risk averse, ex-ante identical students (i, j) exert costly effort to study, learn and obtain a grade in academic tests or exams. Students value high academic performance. For simplicity, we assume that all tests are identical and we exclude the possibility that good performance in one exam affects later performance.[4] Students are hit by idiosyncratic shocks to their endowment of time or to their productivity. Effort is chosen conditional on such shocks, an assumption that, combined with the lack of interlinks across exams, implies that the choice of effort is completely static.

Additionally, agents are allowed to interact according to three different mechanisms: (i) decentralization (Section 2.1); (ii) joint production (Section 2.2); (iii) mutual (full) insurance (Section 2.3). The choice of the mode of interaction is made before observing the realizations of the shocks on the basis of a standard von Neumann-Morgenstern expected utility function. We further assume that active interactions, either in the form of joint production or mutual insurance, are costly and that such costs decrease with how well students know each other.[5] The choice of the mode of interaction can be revised for each exam, for example when the lectures begin.

Utility function. The utility of the generic student i depends positively on her academic performance x_i and negatively on effort e_i. Moreover, we assume two additional properties of the utility function. First, students are risk averse, a property that generates the desire to insure against fluctuations in academic performance. Second, the individual returns to x_i (might) decrease with one's peer performance x_j. Alternative interpretations of this assumption are that students are averse to equality or feel the peers' pressure. We parameterize this relative utility effect with a loading factor $\gamma \in (0, \overline{\gamma})$ that multiplies x_j in i's utility and vice versa. Thanks to this assumption the model delivers within group correlation of the outcomes, even in the decentralized equilibrium. Eventually, we will work with the following static utility function:[6]

$$U_i = \ln(x_i - \gamma x_j) - e_i \tag{1}$$

Production/Learning function. Academic performance is the output of a learning/production process, whose inputs are effort (possibly of both agents) and time t_i, combined according to the following technology:

$$x_i = t_i g(e_i, e_j) \tag{2}$$

[4] The model could easily accommodate interlinks across academic tests and the qualitative implications would remain unchanged.

[5] The term *active interaction* is merely used for expositional clarity and convenience and it should not be given any further interpretation. In the context of this paper it simply refers to interactions that require some cost to be undertaken. The only crucial distinctive feature of the interactions taking place in the decentralization setting is their being costless.

[6] There is a corresponding and symmetric function for agent j. Obviously, γ is bounded above by $\overline{\gamma}$ to guarantee that $x_i > \gamma x_j$ and $x_j > \gamma x_i$. The choice of the particular (concave) utility function and the static framework are not crucial for the testable implications produced.

While the level of effort is endogenously chosen by the student, time is an exogenous factor subject to idiosyncratic shocks.[7] In this interpretation, it is natural to assume that the two production factors are complementary to some degree (not perfect substitutes).[8] We assume:

$$\frac{\partial g(e_i, e_j)}{\partial e_i} > 0, \quad \frac{\partial g(e_i, e_j)}{\partial e_j} \geq 0 \quad \textit{(production is non-decreasing in efforts)}; \tag{3}$$

$$\frac{\partial^2 g(e_i, e_j)}{\partial^2 e_i} < 0, \quad \frac{\partial^2 g(e_i, e_j)}{\partial^2 e_j} \leq 0 \quad \textit{(production is non-convex in efforts)}; \tag{4}$$

$$\frac{\partial^2 g(e_i, e_j)}{\partial e_i \partial e_j} > 0, \quad \textit{(positive cross-partial derivatives or complementarity of efforts)}; \tag{5}$$

$$\frac{\partial g(e_i, e_j)}{\partial e_i} > \frac{\partial g(e_i, e_j)}{\partial e_j}, \quad \textit{(marginal product of own effort larger than that of peer)}. \tag{6}$$

When students do not produce jointly, e_j does not enter the production function for x_i and for notational convenience we label $g(e_i, 0)$ as $f(e_i)$:

$$x_i = t_i g(e_i, 0) = t_i f(e_i) \tag{7}$$

Shocks. Time is equal to a fixed (maximum) endowment normalized to 1 and we allow for a negative shock of size ϵ to arrive with probability $\frac{1}{2}$ and $t_i = 1 + u_i$. The shock hits only one person in the $\{i, j\}$ couple, so that with probability $\frac{1}{2}$, we have $u_i = 0$ and $u_j = -\epsilon$ and with probability $\frac{1}{2}$ we have $u_i = -\epsilon$ and $u_j = 0$. This is equivalent to assuming that the shocks to t are perfectly negatively correlated between the two agents. This stochastic endowment process could be easily generalized and the results would remain qualitatively unchanged as long as some degree of idiosyncratic variation is maintained. The most intuitive interpretation of the shock is absence due to illness or random distraction in the classroom. Although one could alternatively interpret t as some simple (stochastic) productivity parameter, we find our previous interpretation more convenient for expositional purposes.

2.1 Decentralized solution

When students act in isolation they maximize their utility taking each other's behavior as given, as in a simultaneous move game. Even in this simple version, our model still has the potential to generate peer effects through relative utility (or equality aversion).

Without loss of generality, we will assume throughout the paper that agent i has a positive shock ($t_i = 1$) and j has a negative one ($t_j = 1 - \epsilon$):

$$U_i = ln\left[f(e_i) - \gamma(1-\epsilon) f(e_j)\right] - e_i \tag{8}$$

$$U_j = ln\left[(1-\epsilon) f(e_j) - \gamma f(e_i)\right] - e_j \tag{9}$$

[7] The production function of human capital might clearly include other inputs, such as teachers, resources, class size and class composition, to name just a few. However, to keep the model simple and coherent with the empirical application, we focus solely on effort and time. Adding additional inputs that are controlled by the education institution would only complicate the notation, while in the empirical analysis we can fully control for all these additional factors.

[8] We could also work with a more complicated function $g(e_i, t_i, e_j, t_j)$, but the derivation becomes more tedious without any real additional insight. This is due to the fact that t_i and t_j are exogenous factors.

Equilibrium effort levels (under decentralization) - $e_a^D, \forall a = i, j$ - are determined according to the following first order conditions:

$$\frac{\partial f(e_i^D)}{\partial e_i} = f(e_i^D) - \gamma(1-\epsilon)f(e_j^D) \tag{10}$$

$$(1-\epsilon)\frac{\partial f(e_j^D)}{\partial e_j} = (1-\epsilon)f(e_j^D) - \gamma f(e_i^D) \tag{11}$$

Notice that equations 10 and 11 characterize a Nash equilibrium where agents move simultaneously.

Concavity in the utility function leads to the following results:

Proposition 1 *(a) The equilibrium effort level is not smaller for the student affected by the negative shock: $e_i^D \leq e_j^D$, with equality holding when the other student outcome does not not appear in the utility function, i.e. the loading parameter (γ) equals 0. (b) The performance/grade of the student hit by a positive shock is larger than that obtained by the student with a negative shock, i.e. $x_i^D > x_j^D$, where x_i^D and x_j^D are the equilibrium outcomes for agent i and j respectively.*

The proofs can be found in Appendix A.

Further, we can show that:

Proposition 2 *(a) The equilibrium levels of effort of both student i and j are increasing in γ: $\frac{de_i^D}{d\gamma} > 0$ and $\frac{de_j^D}{d\gamma} > 0$. (b) The average performance is non-decreasing in γ, i.e. $\frac{\partial E(x^D)}{\partial \gamma} \geq 0$ for any admissible $\gamma \in [0, \bar{\gamma})$. (c) Under the regularity conditions set out in Appendix A, the dispersion in performances is non-increasing in γ, i.e. $\frac{\partial CV(x^D)}{\partial \gamma} \leq 0$ for any admissible $\gamma \in [0, \bar{\gamma})$.*

Proposition 2 implies that, as the strength of the externalities in academic performance (measured by γ) increases, so does the mean outcome of the group $E(x) = \frac{1}{2}(x_i + x_j)$ since everyone exerts higher effort.

Understanding the implications of relative utility on the dispersion of the equilibrium outcomes is slightly more complicated. We start using equations 10 and 11 to derive the following convenient expression for the coefficient of variation of the outcomes, our measure of dispersion:

$$CV(x) = \frac{SD(x)}{E(x)} = \frac{x_i - x_j}{x_i + x_j} \tag{12}$$

As a first approximation, note that when $\gamma \to 0$, $e_i^D \to e_j^D$ and $CV(x^D) \to \frac{\epsilon}{1-\epsilon} > 0$. As $\gamma \to 1$, $CV(x^D) \to 0$. Thus, the more students care about their relative performance the lower the dispersion in the outcomes. Intuitively, this result is generated by the concavity of the utility function by which student j, who is hit by the negative shock, is placed on a steeper segment of the utility function and, thus, cares more about performance. In fact, under some additional assumptions about the production function $f(\cdot)$, this result generalizes to the entire range of feasible values of γ. In Appendix A we describe such additional assumptions in detail.[9]

[9] A sufficient condition for $\frac{\partial CV(x)}{\partial \gamma} < 0$ for all admissible values of γ is that the third derivative $\frac{\partial^3 f(e_a)}{\partial^3 e_a}, \forall a = i, j$ be positive and sufficiently large in absolute value.

Proposition 2 shows that our simple model produces what are commonly termed peer effects also in the decentralized equilibrium, simply by virtue of relative utility. We will maintain this case as the baseline scenario.

2.2 Joint production

In this section we modify the model under the assumption that students work together in the production process.[10] This could also be interpreted as a model of co-authorship, where each agent individually goes up for tenure or, alternatively, as a model of teamwork where individual performance is verifiable, effort is observed and the agents contribute to each other's output (Alchian & Desmetz 1972, Holmstrom 1982, Kandel & Lazear 1992).

More formally, the model in this section characterizes the Nash equilibrium of a simultaneous move game in which agents choose effort to maximise the utility function in equation 1 and the production technology is defined in equation 2.

Combining the definition of the production function in equation 2 and the structure of the shocks, we can define the following utility functions for agent i and agent j:

$$U_i = ln\left[g(e_i, e_j) - \gamma(1-\epsilon)g(e_j, e_i)\right] - e_i \tag{13}$$

$$U_j = ln\left[(1-\epsilon)g(e_j, e_i) - \gamma g(e_i, e_j)\right] - e_j \tag{14}$$

Like in the decentralization case, equilibrium effort is determined after observing the shocks according to the following first order conditions for the maximization of equations 13 and 14:

$$\frac{\partial g(e_i^C, e_j^C)}{\partial e_i} = g(e_i^C, e_j^C) - \gamma(1-\epsilon)g(e_j^C, e_i^C) \tag{15}$$

$$(1-\epsilon)\frac{\partial g(e_j^C, e_i^C)}{\partial e_j} = (1-\epsilon)g(e_j^C, e_i^C) - \gamma g(e_i^C, e_j^C) \tag{16}$$

where e_i^C is the equilibrium effort of agent i under joint production and similarly for e_j^C.

Using equations 15 and 16, it is relatively easy to derive the following proposition, where e_i^D and e_j^D are the equilibrium effort levels under decentralization:

Proposition 3 *(a) The level of effort under joint production is larger than in the decentralized case for both students, i.e.* $e_a^C > e_a^D$ *for any* $a = \{i, j\}$*. (b) Average performance in the joint production scenario is larger than in decentralization:* $E(x^C) > E(x^D)$*, where* $x^C = \left[g(e_i^C, e_j^C), (1-\epsilon)g(e_j^C, e_i^C)\right]$ *and* $x^D = \left[f(e_i^D), (1-\epsilon)f(e_j^D)\right]$ *are the equilibrium outcomes under joint production and decentralization, respectively.*

The proof, in Appendix A, is trivial given that the level of effort is larger under joint production than decentralization for both students, a result that rests exclusively on the existence of complementarities in the production function (i.e. $\frac{\partial^2 g(e_i, e_j)}{\partial e_i \partial e_j} > 0$).

[10] The term *cooperation* might also suggest a situation where agents maximize the sum of each other's utilities. This is not how we solve the model in this section, although in the On-line Appendix we consider precisely that type of equilibrium as the planner's solution.

Another interesting result of this model is described in the next proposition:

Proposition 4 *The difference of the performances is smaller under joint production than decentralization, i.e.* $(x_i^C - x_j^C) < (x_i^D - x_j^D)$*, hence, given Proposition* $3.a$ *and* $3.b$*, both the standard deviation and the coefficient of variation are smaller under joint production than decentralization, i.e.* $SD(x^C) < SD(x^D)$ *and* $CV(x^C) < CV(x^D)$*.*

The formal proof is in Appendix A but the intuition of the result in proposition 4 is straightforward: the concavity in the utility function implies that the agent hit by the negative shock (j) enjoys higher marginal utility from improved outcomes, hence, given the complementarity in the production function, she increases effort (relative to the decentralized equilibrium) more than her peer.

In sum, this section shows that, under joint production, students' outcomes are higher on average and less dispersed than under decentralization.

2.3 Insurance

The third and last mechanism of interaction that we consider is mutual insurance against idiosyncratic shocks to time or productivity, e.g. health shocks or random distractions (John H. Cochrane 1991, Barbara J. Mace 1991, Robert M. Townsend 1994). Under such circumstances the student would find herself in need of help to fill the gap of important teaching material. Given risk aversion, students have a desire to insure against such fluctuations in their time endowments (or their productivities) by exchanging notes or explanations of class material with classmates. We model such an exchange as a transfer of time or productivity t between agents.

As already mentioned, we choose an extremely simple structure of the shocks but the implications of the model are robust to alternative assumptions. As long as there is some idiosyncratic variation in t, the main results are unchanged.

For simplicity, we consider only the case of full insurance, where the students are perfectly able to smooth away the shocks to their time endowments. Moreover, given that our agents are ex-ante identical, we set identical Pareto weights. For the moment, we also abstract from any issue of commitment (Stephen Coate & Martin Ravallion 1993, Narayana R. Kocherlakota 1996, Fernando Alvarez & Urban J. Jermann 2000, Ethan Ligon, Jonathan P. Thomas & Tim Worrall 2000) but we will come back to this later in section 2.4.

Under these assumptions, agents in our model choose their effort levels exclusively on the basis of the aggregate time endowment in the group. Hence, students solve the optimization problem by simply splitting equally the difference in their endowments. Moreover, since in our model the agents are ex-ante identical, the equilibrium level of effort is the same for both of them: $e_i^I = e_j^I = e^I$. Contrary to the previous cases (decentralization and joint production), in this particular setting it does not matter whether effort is determined before or after observing the shocks, apart from commitment issues. Formally,

$$U_i = U_j = U = ln\,[tf(e) - \gamma tf(e)] - e \tag{17}$$

$$t = (1 - \frac{\epsilon}{2}) \tag{18}$$

The corresponding first order condition is:

$$\frac{\partial f(e^I)}{\partial e^I} = f(e^I) \tag{19}$$

and equilibrium performance is:

$$x^I = (1 - \frac{\epsilon}{2}) f(e^I). \tag{20}$$

Notice that γ, the parameter that measures reference utility considerations, does not appear in these first order conditions. This is due to the fact that under (full) insurance agents are identical both ex-ante and ex-post so that comparisons become meaningless. This has important implications on equilibrium effort, which is necessarily lower under mutual insurance than under decentralization (and, consequently, also compared to joint production) due to the fact that the within-group comparisons that induced higher effort in the benchmark case have now disappeared.

It is also trivial to see that, in this model, the ratio of marginal utilities $\frac{\partial U_i}{\partial e_i} = \frac{\partial U_j}{\partial e_j}$ is equal to 1. This is a consequence of full insurance, as the ratio of the marginal utilities depends only on the Pareto weights, which, in our simple model, are identical for both agents. As such, given standard assumptions on the utility function, the ratio of the levels of performance $\frac{x_i^I}{x_j^I}$ is constant across states (and time).

This observation has a direct testable prediction: the vectors of performance outcomes of any two agents who engage in mutual full insurance should never cross, given that individual performance depends exclusively on the aggregate time endowment and the relative importance of the student in the group (the Pareto weights).[11] In Section 4, we take also this prediction to the empirical test.

We can, then, prove the following, where e^I is the equilibrium effort level, common to both agents, under full insurance:

[11] Such prediction would be however incorrect if the agents had heterogenous risk preferences (Sam Schulloffer-Wohl 2008, Maurizio Mazzocco & Shiv Saini 2012).

Proposition 5 *(a) The levels of effort and performance are lowest in the full insurance scenario for both students, i.e.* $e_a^I < e_a^D < e_a^C$ *for any* $a = \{i, j\}$*, which immediately implies* $E(x^C) > E(x^D) > E(x^I)$*.*
(b) Performances dispersion is lowest under full insurance, i.e. $CV(x^I) < CV(x^C) < CV(x^D)$*.*

The proof is fairly straightforward and can, once again, be found in Appendix A. The intuition is that, under full insurance, effort is lower due to the lack of within-group heterogeneity that, with a reference based utility, generates the desire to excel. The result on dispersion is the combined outcome of the reduced impact of the idiosyncratic shocks, mechanically generated by the insurance process, and the equalization of efforts among the agents. As we show in Appendix A, the variance is reduced at a faster rate than the mean performance, indeed in our simplified model with ex-ante homogeneous agents the (cross-sectional) variance of performance is equal to zero.

2.4 Transition or the choice of the mechanism

In this section we discuss the choice of the mode of interaction as a possible mechanism of transition between the benchmark scenario of decentralization and the two proposed models of joint production and insurance. The mechanics of the transition will form the basis of our empirical exercise, as we will relate the equilibria defined in the previous sections to an underlying exogenous variation in the time students spend together. We expect students not to interact actively when they only met a few times and, as they spend more and more time together, they may start engaging in joint production or mutual insurance (or both).

The choice of the interaction modes between two students is based on a standard expected utility function defined over the two possible realizations of the shocks. Notice that, since agents are ex-ante identical, they both share the same expected utility and the resulting choice of the mechanism will be unanimous.

We also assume that active social interactions, such as those required to produce jointly or to engage in mutual insurance, impose some utility cost and that such costs decrease with how well agents know each other. First, consider joint production and assume that the ability to produce jointly depends on how well the students know each other, which is a positive function of the time they spend together. We model this as a (utility) cost of the following form $c(m) \geq 0$, where m is the number of times agents meet randomly and with $\frac{\partial c(m)}{\partial m} < 0$. This cost function incorporates the idea that agents must pay some initial cost for getting to know each other. At some point the number of previous meetings is large enough to make the cost of cooperating lower than its benefits and students would, then, transit from the baseline scenario of complete decentralization to the joint production state.

A similar argument applies to insurance, as it takes time to get to know someone to the point that reciprocal monitoring and trust are sufficient to engage in mutual insurance. More formally, we describe the cost of insurance with a function $h(m)$, with $\frac{\partial h(m)}{\partial m} < 0$. The function $h(m)$ could also be interpreted as a punishment or sanction function that describes the utility loss derived from deviating from the insurance scheme (Coate & Ravallion 1993). Under such interpretation, the size of the punishment increases with the number of meetings, somehow capturing the idea that defaulting on someone we know better is more painful or more difficult.

Given that our agents are risk averse, they will have an incentive to renege on the insurance contract in the good state ($t = 1$) and, as such, the contract will have to be self enforcing or incentive compatible (Coate & Ravallion 1993, Ligon, Thomas & Worrall 2000). However, given the static nature of this simplified model, the only equilibrium would be to renege if the sanction is low. As students meet more often, they either find it harder to default on good friends or they develop the ability to punish deviating behaviors more effectively, so that, even under limited commitment, the full insurance allocation can be implemented for a sufficiently high level of m.

Consistently with our empirical application, where students initially do not know each other and are randomly allocated to teaching classes, we assume that decentralization requires no costs and is the baseline scenario.

Students select the mechanism that is associated with the highest level of expected utility as follows:

$$\begin{aligned} E(U^j) &= \mathbf{1}\{j=D\}\left(\frac{1}{2}U(e^D_{u=0})+\frac{1}{2}U(e^D_{u=-\epsilon})\right) \\ &+ \mathbf{1}\{j=C\}\left(\frac{1}{2}U(e^C_{u=0})+\frac{1}{2}U(e^C_{u=-\epsilon})-c(m)\right) \\ &+ \mathbf{1}\{j=I\}\left(U(e^I)-h(m)\right), \quad j=D,C,I. \end{aligned} \tag{21}$$

Where $\mathbf{1}\{.\}$ is the indicator function, $U(e^D_{u=0})$ is the utility computed at the equilibrium effort level in decentralization under the good realization of the shock and similarly for the other expressions, with superscripts indicating the interaction mode and subscripts the values of the shock. Recall than in the mutual insurance model equilibrium effort is independent of the shock.

Agents choose the mode of interaction J in order to maximize equation 21. When m is sufficiently low, possibly zero, the costs associated with joint production or mutual insurance are larger than the benefits and students simply act in isolation according to our benchmark model of decentralized behavior. As m increases, both $c(m)$ and $h(m)$ decrease until, for a sufficiently high m, either joint production or mutual insurance become the utility-maximizing mode of interaction.

The relative shape of the functions $c(m)$ and $h(m)$ defines which of the models, joint production or insurance, arises first. In fact, although we do not model this situation explicitly, it might very well be possible that, for some values of m, both joint production and mutual insurance coexist. Moreover, while it is obvious that in the absence of costs both joint production and mutual insurance would be preferred to decentralization, preferences over these last two mechanisms depend on the functional forms as well as on the size of the shock.[12]

Ultimately, the type of mechanism at work is an empirical matter and our analysis in Section 4 exploits the following testable implications to shed light on this precise issue:

$$\begin{aligned} &\text{If} \quad \frac{\partial CV(x)}{\partial m}<0 \quad \text{and} \quad \frac{\partial E(x)}{\partial m}>0 \quad \Rightarrow \quad \text{Joint production is the prevailing mechanism;} \\ &\text{If} \quad \frac{\partial CV(x)}{\partial m}<0 \quad \text{and} \quad \frac{\partial E(x)}{\partial m}<0 \quad \Rightarrow \quad \text{Insurance is the prevailing mechanism.} \end{aligned}$$

[12]Joint production offers the advantage of exploiting the complementarities in the production function, hence in the absence of costs it would be preferred to decentralization. Similarly, mutual insurance reduces the variance of the outcomes and, given the concavity of the utility function, it dominates decentralization when $h(m) = 0$.

The reduction in the (within-group) dispersion of performances as a function of the time spent together indicates the existence of peer effects, while the relation between the number of meetings and average performance in the group allows us to distinguish between joint production and insurance.

Notice additionally that these empirical implication have been derived under the assumption that the intensity of interactions m only affects the costs of engaging in joint production or insurance. In particular, we do not allow m to affect the form of the utility function and namely the degree to which agents care about each other's performance (γ). We return to this point in Section 5.

The model in this section rests on a number of simplifying assumptions, many of which can be relaxed without affecting the main implications. In particular, in the On-line Appendix we extend the model to a setting with multiple agents and we derive the social planner solution.

3 The data

In this section we describe the data we use in Section 4 to test the theoretical predictions of our models. The data come from the administrative archives of Bocconi University, an institution of higher education located in Milan, Italy (De Giorgi, Pellizzari & Redaelli 2010, Giacomo De Giorgi, Michele Pellizzari & William Gui Woolston 2012). The important feature of these data for the purpose of our application is that students are (repeatedly) randomly assigned to teaching classes for each of their compulsory courses.

More specifically, we will focus on two cohorts of students who first enrolled at Bocconi in the academic year 1999/00 and 2000/01.[13] These students were offered 7 degree programs (majors), however, only 3 of them were large enough to require the splitting of lectures into more than one class: *Economics*, *Management*, *Economics and Finance*.[14] The official duration of all programs was 4 years and, during the first two years, and for most of the third, all students were required to take a fixed sequence of compulsory courses. Afterwards students could choose elective subjects, within some program-specific rules. In order to avoid issues of endogenous selection, we exclude all elective courses from our analysis ad we focus exclusively on compulsory courses.

The compulsory curricula is summarized in Table 1, while more details are given in the On-line Appendix. The table reports the number of compulsory courses (columns 1, 3 and 5) and the total number of lecturing hours (columns 2, 4 and 6) for each degree program and academic year. For example, Table 1 shows that students in Management take 8 compulsory courses during their first year for a total of 464 hours of lectures.

[TABLE 1]

The crucial institutional feature for our testing strategy is the completely random procedure adopted

[13] Although we have access to records for all students enrolled since 1989 and until 2009, we cannot choose a cohort that is too recent because some of the students might still be working towards the completion of their degrees. For the earlier cohorts we do not have information on the class identifiers, which are essential for our analysis. Moreover, Bocconi reformed the structure of its programs twice during the period covered by these data, first in 1999/00 and then 2001/02. Hence, to avoid comparing cohorts across different systems, we consider only students enrolled in their first year in 1999/00 and 2000/01.

[14] The other programs were *Economics and Management of the Public Administration*, *Economics and Law*, *Law*, *Economics and Management in Arts, Culture and Communication*.

to allocate students to teaching classes.[15] The randomization was repeated at the beginning of each academic year, when each student was informed of her class identifier and was instructed to take all the lectures of the year in the class corresponding to her assigned identifier. Namely, students would take all the courses of the first year with the same random group of peers, then all the courses of the second year with a different random group and so on. Elective courses, which we do not consider in our analysis, were usually much smaller in size and could easily be taught in a single class.

The figures reported in Table 1 clearly show that students who were assigned to the same class ended up physically sitting in the same classroom for a considerable amount of time, thus suggesting that the random allocation process is an important determinant of the strength of students' interactions.

Students were allocated into several classes for the explicit purpose of maintaining adequate class sizes and to allow teachers to interact with the students in a more direct way. The yearly repetition of the random allocation was, instead, justified with the desire to encourage interactions among students. Bocconi has followed attentively the rule of randomly allocating students to teaching classes so as to avoid clustering of students in some classes. Moreover, for organizational reasons, students allocated to a specific class were also taking most or all of their courses in exactly the same classroom. Overall, there are 12 classes per academic year (and cohort): 8 classes in Management, 2 in Economics and 2 in Economics&Finance. The classes are approximately equal sized, although not exactly.[16]

[TABLE 2]

Table 2 reports some descriptive statistics on the 2,406 students that we eventually consider in our empirical exercise, broken down by degree program. The large majority of students are enrolled in the Management program, which attracts over 70% of them. The Economics program is chosen by about 10% of the students and the remaining 20% are in Economics&Finance. There is a higher incidence of males (58.5%) than females (41.5%) in the student body, with gender differences being more pronounced in the Economics and Economics&Finance programs. The majority (68%) of students come from outside the province of Milan, the site of Bocconi. We also have information on the students' household income, which is recorded in 4 brackets at the time of enrollment for the purpose of determining tuition fees and eligibility for scholarships. About 22% of the students are in the highest of those brackets (approximately above 140 thousands USD). Average admission test scores and high school grades, normalized on a scale 0-100, suggest that the best students cluster in Economics and Economics&Finance.

[FIGURE 1]

In Figure 1 we present evidence consistent with random allocation, as in De Giorgi, Pellizzari & Woolston (2012) and De Giorgi, Pellizzari & Redaelli (2010). The figure compares the distribution of

[15] The terms *class* and *lecture* often have different meanings in different countries and sometimes also in different schools within the same country. In most British universities, for example, lecture indicates a teaching session where an instructor - typically a full faculty member - presents the main material of the course. Classes are instead practical sessions where a teacher assistant solves problem sets and applied exercises with the students. At Bocconi there was no such distinction, meaning that the same randomly allocated groups were kept for both regular lectures and applied classes. Hence, in the remainder of the paper we use the two terms interchangeably.

[16] As discussed in De Giorgi, Pellizzari & Woolston (2012), variation in class size is generated by logistic constraints, i.e. variation in the physical size of available classrooms.

some selected characteristics in the entire cohort and within the groups of peers of a randomly selected student in each of the three initial academic years. Specifically, the variables that we consider for the tests in Figure 1 are the entry test score (upper left graph), high school grade (upper right graph), the gender indicator (lower left graph), the indicator for high income (lower middle graph) and the dummy for residence outside Milan (lower right graph). As it is evident from the figure, the distributions all look very similar. Statistical tests also confirm such a visual impression. In the On-line Appendix we also report Kolmogorov-Smirnov tests for the comparison of the distributions of the admission test scores in any possible pair of classes and any possible academic year and results show that we reject the null of equality of the distributions, at the 10% level, in fewer than 10% of the comparisons (11 out of 180), which is what one should expect under random assignment. We have produced similar tests also for the distribution of high school grades as well as all the other characteristics considered in Figure 1, using tests of proportions for those characteristics that take the form of simple dummy indicators (gender, high income students and residence outside Milan). Results are not reported for brevity but are available from the authors upon request.

In the empirical analysis of Section 4, we consider all possible pairs of students and compare mean grades and coefficients of variation (as a measure of dispersion of outcomes within the pair), both across pairs who are randomly allocated to the same classes more or less often and for the same pair over academic years.

In principle, we could conduct the same analysis at the level of the individual student by defining individual-specific peer groups identified by students who were randomly allocated to one's same class in any given academic year and/or in the past.[17] There are both practical and theoretical reasons why we prefer the pairwise specification to this alternative approach. First, the focus on student pairs allows us to directly compare our results with the implications of the model, that is also presented in a two-agents setting. Second, there is no obvious definition of the appropriate measure of meeting times at the level of the group. One could compute the number of times all students in the group have been simultaneously allocated to the same class, an event that is relatively rare given the large size of most classes. Alternatively, one could consider the mean of the pairwise meetings, but this approach would merely aggregate information that can be more effectively used in the pairwise setting. Ideally, the choice should be guided by some presumption about the level at which interactions take place, either one-to-one or within groups. In the absence of direct information on this issue, the pairwise approach provides the most stringent test of the implications of our theoretical model. In fact, while the student-level analysis may easily miss out on interactions that take place one-to-one, the opposite is true only under relatively strong assumptions about the form of interactions, such as the case in which interactions exclusively take place in groups. The On-line Appendix further discusses how our pairwise approach can be reconciled also within a model of multiple agents interacting in groups.

Practically, we construct all pairs of students who could possibly meet, i.e. all pairs of students who are in the same degree program and enrollment cohort. Eventually, we construct a dataset of over 800,000 pairs, whose descriptive statistics are reported in Table 3. It is important to clarify that the data used for the empirical analysis contain one observation per pair per period, so that if the pair i, j appears

[17] In fact, in a previous version of the paper we did adopt this approach obtaining similar results to those reported here.

in the data the symmetric pair j, i does not. Specifically, since the random allocation is performed within degree program and cohort, only students within degree program and cohort cells can be allocated to the same class. Call N_{cp} the number of students in cohort c and degree program p, then the number of non-symmetric pairs generated by those students is $\frac{N_{cd}(N_{cd}-1)}{2}$.[18]

[TABLE 3]

All the descriptive statistics in Table 3 are consistent with the random allocation (within degree programs and cohorts) and the descriptives of the student body from Table 2.[19] The last row of Table 3 reports the mean value of a dummy indicator for whether the students in the pair are ever allocated to the same class and it shows that slightly less that 40% of the pairs ever met. Given the size of the degree programs and the number of random classes in each of them, this number varies considerably with a low 33% in Management and a much larger 87% in Economics and Economics&Finance.

The repeated randomization adopted by Bocconi generates as a byproduct the fact that some students met more often than others, a crucial feature for our empirical exercise in Section 4. In Table 4 we describe the main variables that arise from such a repeated randomization process together with our key outcomes at the level of the student pairs.

[TABLE 4]

The first row of Table 4 reports the average number of courses any two students sit together by degree program and academic year. It is important to emphasize that, since this is meant to be an indicator of how many opportunities for interactions students have had, it is a cumulative variable, i.e. it measures the number of courses together in the current and in the past years. This is the reason why the number of courses together always increases from one year to the next.[20] On average two random students would have sat 1.2 courses together by the end of their first year, 2.3 by their second year and 3.5 by the third. Notice that these averages combine pairs of students who never meet and others that meet repeatedly. Given the number of classes and courses in the different majors the above statistics become larger in Economics and in Economics&Finance, i.e. 10-11 courses by the third year.

Since each course is taught for a different number of hours, the number of times students in the pair are randomly assigned to the same class defines the number of hours they end up sitting together in the same classroom (in the current and in the previous years). The average students pair spends approximately 70 hours together by the end of the first year, 135 by the second and over 200 by the third. Given the variation in the number of courses and their duration in hours (see Table 1), these statistics vary considerably across degree programs, with students in Economics and Economics&Finance enjoying a lot more opportunities for interactions.

We also describe the evolution of both our main outcomes: the average and the dispersion of academic performance within groups and across academic years. Notice that our main explanatory variable

[18]Table Web.3 in the On-line Appendix shows the number of students in each cd cell and the corresponding number of pairs.

[19]The comparison of Table 2 and Table 3, in fact, can also be interpreted as a randomization test. One could derive most of the statistics in Table 3 from those in Table 2, using the number of classes in each program and the size of each enrollment cohort (in each degree program).

[20]In the regression analysis of Section 4 this mechanical time effect is controlled for by a set of year and degree program dummies, which we always include in the control set.

(i.e. hours together) varies only across academic years for the same pair of student, hence we consider only one observation for each pair in each year. Consequently, we compute the average grade of the pair over all the compulsory courses of each academic year. As a measure of dispersion we consider the coefficient of variation.[21] Table 4 shows the basic characteristics of these variables, broken down by academic year and degree program. The average grade is similar across all programs and years and fluctuates around a mean of approximately 26/30 (B+), with a coefficient of variation of about 0.1.[22]

The last row of Table 4 reports the means, by degree program and academic year, of the crossing indicator that we use in Section 4 to test the crossing property (see Section 2.3). Such an indicator is constructed by comparing the grade vectors of the two students in each pair in each academic year and verifying whether such vectors cross (at any point). In other words, the crossing indicator is equal to zero if one student consistently outperforms (or ties) the other in all the courses of the specific academic year. Otherwise, the grade vectors cross and the indicator takes value 1. The probability of crossing, for any two students, ranges between 80% and 88%, with the exception of the third year in the Economics program where it is equal to 68%.

For completeness, in Table 5 we present the full extent of variation in our key independent variable, i.e. number of meetings or hours together. Given the academic structure of the curricula, the variation arises both between and within (over-time) programs. There are a number of sources of variation for our independent variable, all arising from the repeated random allocation in different teaching classes: variation across majors in a given year, as well as variation between pairs (cross-sectional and over time) and within pairs overtime. In particular, in the Management major about 65% of the possible pairs never meet, while those pairs who meet in the first and second year (1.4%) sit together for about 850 hours, the few who meet every year (0.2%) end up spending over 1,300 hours together. Any two students in Economics are quite likely to meet as there are only two classes. In fact only about 13% of the (potential) pairs never meet, while those who meet in year 1 and 2 (12%) spend 880 hours in the same classroom, those who always meet (12%) spend about 1,100 hours together. Consistently with the random allocation mechanism and similar class sizes, the meeting probabilities in the Economics&Finance program are very similar to those in Economics. One can also show that as a result of the random allocation in two classes of similar sizes (Econ and Econ&Finance) the distribution of meetings, i.e. the leftmost column in Table 5 should look pretty even and close to 12-13%, which indeed is what we find in the data, an additional piece of evidence that is consistent with a successful random allocation.

[TABLE 5]

[21] In computing both the mean grade and the coefficient of variation we weight courses by their length, assigning a higher weight to longer courses, i.e. courses with more hours of lectures. For robustness, we have replicated all our results using unweighted outcomes obtaining virtually identical estimates. The results are not reported for brevity but are available from the authors upon request.

[22] In Italy, university exams are graded on a scale 0 to 30, with pass equal to 18. Such a peculiar grading scale comes from historical legacy: while in primary, middle and high school students were graded by one teacher per subject on a scale 0 to 10 (pass equal to 6), at university each exam was supposed to be evaluated by a commission of three professors, each grading on the same 0-10 scale, the final mark being the sum of the three. Hence, 18 is pass and 30 is full marks. Apart from the scaling, the actual grading at Bocconi is performed as in the average US or UK university.

4 Empirical Analysis

In this section we analyze how the mean and the dispersion of academic outcomes within groups evolve as students spend more and more time together over the first three years of their academic tracks. Following our theoretical discussion in Section 2, the different mechanisms of social interactions have different implications for the relationships between number of meetings (hours) and the mean and dispersion of performance. Our main testable predictions are the following:

$$\begin{array}{llllll}\text{If} & \dfrac{\partial CV(x)}{\partial m} < 0 & \text{and} & \dfrac{\partial E(x)}{\partial m} > 0 & \Rightarrow & \text{Joint production is the mechanism;}\\ \text{If} & \dfrac{\partial CV(x)}{\partial m} < 0 & \text{and} & \dfrac{\partial E(x)}{\partial m} < 0 & \Rightarrow & \text{Insurance is the mechanism.}\end{array}$$

where m is a measure of the number of hours the pair of students spend together. Namely, as the number of interactions increases, we expect the average grade to increase if joint production is the prevailing mechanism, to decrease if it is insurance and to stay unchanged if the leading model is decentralization. The implications for dispersion are less clear cut, as we simply expect the dispersion of grades within pairs to decrease as the number of meeting or hours spent together increases when the prevailing mechanism is either joint production or insurance (or any combination of the two) and to remain unaffected if no transition takes place from decentralization to any of the other models.[23]

We test these predictions in Table 6, where we report the results of the following regressions:

$$y_{ijcpt} = \alpha_1 + \beta_1 m_{ijt} + \gamma_{1t} + \delta_{1p} + \zeta_{1c} + u_{ijcpt} \quad (22)$$

$$y_{ijcpt} = \alpha_2 + \beta_2 m_{ijt} + \gamma_{2t} + \eta_{ij} + e_{ijcpt} \quad (23)$$

where y_{ijcpt} is, alternatively, the mean grade or the coefficient of variation or the crossing indicator for pair (i, j), in cohort c, degree program p in year t; m_{ijt} measures the number of meetings or hours student i and j spent together on or before year t. Additionally, we control for year, degree program and cohort effects: γ_t, δ_p, ζ_c respectively. Equation 23 is a slightly different specification of equation 22, where we include a pair fixed effect η_{ij}, which obviously also captures the degree program and cohort effects. u_{ijcpt} and v_{ijcpt} are random error terms.

It is important to notice that although we conduct our analysis running linear regressions in m_{ijt} our models of interactions suggest a non-linear data generating process, as cooperation and insurance will occur once the frequency of students interactions crosses a certain threshold. In fact the cross-sectional variation in outcomes between students would arise from such non-linearities within a specific major even if class-sizes and teaching hours were all the same both within and across years. We prefer the linear specification in the empirical analysis as it is able to capture our testable implications in a parsimonious way. Further, a non-linear specification would be quite difficult to interpret given that joint production, insurance and decentralization can coexist in the data.

In addition, detecting non-linearities or threshold effects in our data is particularly complicated for at least two reasons. First, from the empirical point of view, we do not observe data over the entire

[23] If we allow γ to be increasing in m, i.e. the relative importance of one's peer performance increases in the number of meetings, we have that $\frac{\partial E(x^D)}{\partial m} > 0$ and $\frac{\partial Var(x^D)}{\partial m} < 0$, as shown in Proposition 2.

distribution of hours together. After pairs who have never been assigned to the same class, pairs who have been assigned together only once already spend around 400 hours in the same classroom. Hence, we cannot detect any non-linear effect between zero and 400 hours. Secondly, from the theoretical standpoint, if the thresholds beyond which students start interacting were heterogeneous across pairs and possibly courses, as it would be reasonable to assume, the possibility to statistically detect any non-linearity in our main specifications would depend crucially on the distribution of such thresholds along the distribution of hours together. In fact, in the presence of heterogeneous thresholds non-linearities may completely disappear.

The computation of the correct variance-covariance matrices for the estimators of the parameters in equations 22 and 23 is particularly complicated, as our data feature several non-standard characteristics: (i) the panel dimension (recall that we use one observation for each pair in each academic year); (ii) the pairwise structure of the observations; (iii) the semi-aggregate level of variation of our main regressor (hours together, m_{ijt}) and (iv) the network structure.

We take care of the first problem (the panel dimension) in equation 22 by applying the standard random effect transformation and in equation 23 by transforming the model in orthogonal deviations.[24] In the standard textbook panel model, such transformations eliminate any serial correlation as well as any heteroskedasticity from the error terms, so that the classic random or fixed effects estimators can be computed by simple ordinary least squares on the transformed data.

We can then take into account the other issues by adjusting the transformed errors. Since such an adjustment can theoretically be performed under different sets of assumptions, we compute two alternative versions of the asymptotic z-statistics to support the statistical significance of our estimates.

In the first version we allow the off-diagonal elements of the variance covariance matrix of the transformed error terms of equations 22 and 23 to be correlated within non-nested clusters defined at the level of each individual student (i and j) and at the exact level of variation of the key regressor of interest m_{ijt}. More specifically, we assume that the error term u_{ijcpt} (e_{ijcpt}) is the sum of three components: a student i component that induces serial correlation among all pairs where student i is a member, a student j component that induces serial correlation among all pairs with student j as a member and a third component that is common to all pairs that share the same value of m_{ijt}. As long as these components are all additive, the random- and fixed-effects transformations do not affect the structure of the transformed error terms, which remain additive in the three components with the same (qualitative) variance-covariance matrices.

Hence, we take proper account of the pairwise nature of our data by allowing serial correlation within clusters at the level of each individual student. Additionally, since our main regressor of interest

[24] The standard random effect transformation subtracts from each variable in the model (both the dependent and each of the regressors) its within mean scaled by the factor $\theta = 1 - \sqrt{\frac{\sigma_u}{T\sigma_\eta + \sigma_u}}$. For example, the random-effects transformed dependent variable is $y_{ijcpt} - \theta\overline{y}_{ijcp}$, where $\overline{y}_{ijcp} = T^{-1}\sum_{t=1}^{T} y_{ijcp}$ and $T = 3$ in our application. Similarly for all the regressors. The estimates of σ_η and σ_u that we use for this transformation are the usual Swamy-Arora, also used by the command *xtreg* in Stata (P. A. V. B. Swamy & S. S. Arora 1972). The transformation in orthogonal deviations consists in taking all variables in the model (both the dependent and each of the regressors) in differences from the within mean of all future values, scaling by a factor c_t that guarantees homoskedasticity (in the absence of other error components). For example, the dependent variable of equation 23 transformed in orthogonal deviations is $\left[y_{ijcpt} - (T-t)^{-1}\sum_{k=t+1}^{T} y_{ijcpk}\right] c_t$, where $c_t^2 = \frac{T-t}{T-t+1}$. Notice that in the model in orthogonal deviations the time dimension is reduced by one unit, which explains the different number of observations in columns 2 and 4 of Table 6 compared to columns 1 and 3 (and similarly in Table 7 and Table 9).

(m_{ijt}) varies at a more aggregated level that the student pair (i.e. all pairs whose members are assigned to the same classes in current and previous years are associated to the same value of m_{ijt}) we face a standard Moulton problem (Brent R. Moulton 1990) and we further allow serial correlation along a third dimension that corresponds to the exact cells of variation in m_{ijt} (such cells of variation are those shown in Table 5).

Technically, we perform such a three-level clustering using the procedure described in Colin A. Cameron, Douglas M. Miller & Jonah B. Gelbach (2011), which does not require to specify the form of the correlation within clusters and is very easy to implement practically even with a large sample like ours. However, it does not take into account the network structure of our data.

The second version of the asymptotic z-statistics that we compute does take explicit account of the network by estimating the full variance covariance structure of the error terms under the assumptions spelled out in detail in Appendix B. In practice we allow for flexible parametric correlation between the elements of the variance covariance matrix. Given the high computational burden of the matrices involved in such a computation, we construct those statistics on a 10% random sample of our students.[25] Notice that we random draw a sample of student in the first year and construct the full network of those students so that there is no measurement error in the network structure.[26]

The identification of the parameters in equations 22 and 23 is straightforward, given that the right hand side variable of interest (m_{ijt}) is generated exclusively by the random allocation process, hence it is fully exogenous. Consistently with this interpretation, we obtain very similar estimates across the two specifications, although including the pair fixed effect η_{ij} substantially improves efficiency.

Obviously, our measure of the strength of social interactions (m_{ijt}) does not necessarily represent the true number of hours any two students spend together, either studying or engaging in other social activities. If we were able to observe such true measure of time together, we could use it in equations 22 and 23 instead of m_{ijt}. However, such a measure would be endogenous, as people may choose to spend more time with some classmates for reasons that are connected to academic performance. Hence, we could use m_{ijt} as an instrument for real time together. Then, our results can be interpreted as the reduced form estimates of this hypothetical IV framework.

Once again, we expect the effect of hours together on the pair's average performance (i.e β_s for $s = 1, 2$) to be positive or negative depending on whether the prevailing mechanism is joint production or insurance, respectively. Our model does not exclude the possibility that both mechanisms operate simultaneously, so that we are really only able to test which one dominates.

[TABLE 6]

The results are reported in Table 6 and Table 7. In columns 1 and 2 of Table 6 we show the estimates of equations 22 and 23 with the mean grade of the pair as a dependent variable. For ease of interpretation

[25]Notice that, given the computational burden, we present both z-statistics only in our main tables 6 and 7. We have experimented with several draws of similar and smaller sizes and the results are qualitatively unchanged, larger draws are not feasible even on very powerful machines (above 100Gb of RAM).

[26]Since the choice of the mechanism is unlikely to be invariant to the structure of the network, our estimates of the parameters in equations 22 and 23 based on the full or the 10% random sample may not be comparable. This is the reasons why, for presentational ease, we prefer to report asymptotic z-statistics rather than standard errors, so as to avoid reporting also the coefficients estimated on the random sample, which are, however, available from the authors upon request.

both the dependent and the explanatory variables in these regressions are normalized within degree program and academic year, so that the estimated effects can be readily interpreted in terms of standard deviations. Namely, a one standard deviation increase in the number of hours together (approximately 240 hours) decreases mean performance by 2.2% to 2.7% of a standard deviation, corresponding to almost 0.05 grade points. It is also interesting to compare the average performance of pairs of students who are always assigned to the same class for all the three years of their track and pairs of students who are never assigned to the same class. These two types of student pairs are approximately 6.3 standard deviations apart in terms of hours together, hence their mean grades are 17% of a standard deviation apart in favor of the pair who never met, corresponding to almost 0.3 of a grade point. Consistently with the random nature of m_{ijt}, including pair fixed effects in the specification of the model only affects the coefficients marginally, while it substantially improves the precision of the estimates.

In columns 3 and 4 of Table 6 we report the estimates of equations 22 and 23 with our measure of grade dispersion as a dependent variable. Results indicate that a one standard deviation increase in the number of hours together (approximately 240 hours) reduces the coefficient of variation of exam grades by 0.7-0.9% of a standard deviation. The comparison of pairs at the extreme ends of the hours distribution indicates that the exam grades of students who are consistently assigned to the same class is approximately 5% of a standard deviation lower compared to those who never sit courses together.

While the negative relationship between dispersion in performance and the number of hours students spend together would be consistent with both insurance and joint production (or a combination of the two), the only mechanism that can rationalize the decline in average performance is insurance.

As discussed in the theoretical section, an additional test to assess the relevance of insurance motives relies on the time-series properties of students' performances, i.e. on whether students' grades in a pair cross each other. The logic of this *crossing property* is the following: consider a pair of students who perfectly share risk by insuring each other (full insurance), then they always exert the same level of effort, hence their outcomes should be either always identical, in the case of homogenous agents (with the same Pareto weights), or never cross if agents are heterogeneous. Crossing outcomes are a signal of less than full insurance. This result has been shown in Cochrane (1991) and Mace (1991) and essentially restates that under full insurance the ratio of marginal utilities between any two agents is always constant and equal to the inverse of the ratio of the Pareto weights. Similar tests are produced in Tullio Jappelli & Luigi Pistaferri (2006) and Mazzocco & Saini (2012) in the consumption literature.

In our application, we expect students to insure more and more as they spend more and more time together, eventually approaching full insurance (see Section 2.4). We test this prediction by estimating equations 22 and 23 with the crossing indicator on the left hand side (linear probability model). The crossing indicator is a simple dummy variable equal to 1 if the vectors of exam grades of student i and j ever cross over the entire series of compulsory courses taken during year t. The results are reported in Table 7.

[TABLE 7]

Results show that the likelihood of crossing decreases with the number of hours spent together and we take this as further evidence consistent with the insurance mechanism. Once again, the estimates are significant at conventional statistical levels only when conditioning on pair fixed effects (column

2). The magnitude of the effect is such that a one standard deviation increase in the number of hours together reduces the likelihood of crossing by 0.4 percentage points or 1.1% of a standard deviation. Alternatively, one can look at the difference between pairs of students who are always in the same class and pairs who are never in the same class, the latter pair would be 2.5 percentage points more likely to display grade-crossing.

4.1 Further Evidence

A first simple robustness check of our results consists in investigating whether, for any spurious reasons, future classmates appear to be related in terms of academic outcomes. If that were to happen, our approach would be falsified from the start, as in none of our model we predict a relation between outcomes of students who have not met yet. In this light, we run a series of placebo regressions where we limit our sample to the first academic year (i.e. one observation for each pair) and we regress our outcomes on an indicator for whether the pair will ever meet in the future two years. Results are reported in Table 8 and, consistently with the idea that future meetings should not be related to current outcomes, we find that all estimates are very close to zero and insignificant.

[TABLE 8]

A second piece of additional evidence addresses the following concern: although the number of hours together arises from the random allocation process, it might still be that our results on dispersion and grade crossing are simply the byproduct of the structure of our data. In fact, pairs with more hours together are more likely than others to have attended courses with the same teachers and, if teachers are heterogenous, this alone might reduce the variance of outcomes within pairs. Similarly, there could be cases in which such a reduction in the variance might also lead to a lower probability of crossing (although one would have to think about some very special cases of teacher heterogeneity). Note that these concerns bear no implications for the mean outcome of the pair as a function of the number of meetings.

To address these issues, in Table 9 we augment equations 22 and 23 with a set of dummies for the classes assigned to each student in the pair (i, j) in every academic year. Overall, there are 12 such dummies (one for each of the 8 classes in Management plus 2 classes in Economics and Economics&Finance) for each student in the pair, for each academic year and each cohort, for a total of 144 dummies. Such class effects are meant to control for the potential bias due to unobservable factors at the class level, such as teacher quality, class size or class composition, that might be partially collinear with m_{ijt}. Results are qualitatively identical to those in Table 6 and Table 7. Also the magnitude of the estimates is very comparable, especially for grade crossing.[27]

[TABLE 9]

The results in the previous section are consistent, within our framework, only with the insurance mechanism. As it is well known in the literature (Alvarez & Jermann 2000), the level of informal

[27] The results are robust to directly controlling for class size in equations 22 and 23.

insurance between two parties under limited commitment rests on three fundamental parameters, i.e. risk aversion, discount rate and correlation of shocks, plus the ability of the parties to monitor and enforce punishment for deviating behaviors. Although we do not directly observe any of these three parameters, one might think that gender is an important determinant of the degree of insurance under the assumption that reciprocal trust and monitoring are easier among same sex pairs.

We explore this idea in Table 10, where we report estimates of equation 23 for samples of pairs restricted to either same gender (row 1) or mixed gender (row 2) or both females (row 3) or both males (row 4). Results are remarkably similar across these groups and it only appears that females have a smaller negative effect on average performance when they interact among themselves.

[TABLE 10]

In Table 11 we further investigate heterogeneity in ability. Although the level of ability of a student in a pair should not matter for the degree of sustainable insurance in a pair, it is however true that it could determine the sharing rule, i.e. the more able of the pair gets allocated more time (or resources, our t parameter in the model) in every state of the world (a higher Pareto weight). Therefore, in order to be consistent with the proposed mechanism, individual ability (which we measure with the standardized entry test) should be irrelevant for the relation between our first and second moments of the outcomes and the exogenous number of hours spent together.

Consistently with this interpretation, in Table 11 we find that, when estimating equation 23 on subsamples defined by ability quartiles of the two students in each pair, results are extremely homogeneous, especially as far as mean outcomes are concerned.

[TABLE 11]

5 Alternative Mechanisms

So far we have considered three plausible mechanisms that are able to explain what is commonly termed peer effects or social interactions in the literature. However, it might still be possible that other mechanisms are consistent with our empirical evidence and, although our results bear important policy implications for the specific setting under consideration, their external validity rests crucially on the theoretical interpretation that we attach to them.

Below we discuss, and for the most part rule out, a series of alternative mechanisms. For example, one could construct a behavioral model where students get to know each other over time and divert their attention from studying to other more socializing activities (going out, partying, etc.). Such a mechanism could easily account for the drop in performance but its implications on the dispersion of outcomes depend crucially on the assumptions about the size of the drop for the high and low performing students. If performance declines more at the top of the distribution than at the bottom, dispersion would be lower, the opposite if the drop is larger at the bottom.

One possible way to adapt such a model to generate lower dispersion in outcomes consists in introducing some kind of preference for equality or conformity (Douglas B. Bernheim 1994, Akerlof 1997), whereby students dislike to perform differently from their friends. In this case students hit by different types of shocks would help each other. This, however, resembles quite closely our insurance model.

One might think as well that such alternative mechanism is particularly valid for lower ability students, as the forgone benefits of studying instead of partying might be lower for this type of students. For this purpose, the evidence presented in the previous Section 4.1 (Table 11) is particularly informative, as we find that the estimated effects are pretty homogeneous across ability combinations. Such a lack of heterogeneity (also across gender) is consistent with insurance, as the insuring parties would not care about their relative ability but focus on the correlation of shocks and the ability to implement the informal contract.

Another possibility is that as students know each other better they care more about their relative performance, in our model this would mean that γ is increasing in m. However, as shown in Proposition 4, if that were the case the mean outcome should increase with the number of meeting, which is the opposite of what we find in the data.

Further, the disruption model of Edward P. Lazear (2001), adequately modified, might be able to produce what we observe in the data. However, the Lazear's model of disruption naturally produces a negative correlation between group size and performance, while our results in Table 9 are not consistent with such a story. In fact when we control for class identifiers, which capture, among other things, the size of the class, we find almost the same results as those in the main Table 6. Moreover, the simplest disruption model would explain why the mean performance falls but would have very little to say regarding the cross-sectional variance, unless some further assumptions are made.

An important alternative model is that of social learning, as in (Enrico Moretti 2011), where the individual consumption of movies is affected by peers' consumption when the quality of the movie is uncertain. In our context this could be the case if students interactions can reduce the noise around the taught material by exchanging information and notes. This seems quite similar to our insurance model at a first glance, however the similarities between the two mechanisms really depend upon the specific learning model one has in mind. For example, an extremely simple model of social learning would essentially translate into a model where the noise in learning, or in our case the size of the shock u, falls as students interact more frequently. In our setup this could be modeled as $t = 1 + \widetilde{u}$, with $\widetilde{u} = u/m$. Where the notation is identical to the one used earlier in the paper. With this formulation it is easy to show that the average (pairwise) performance would increase in the number of meetings since the noise $\widetilde{u}$ would disappear as the number of random meetings increases. In the limit $lim_{m\to M}\widetilde{u} = 0$, so that uncertainty would disappear and the average performance would be, then, a positive function of the number of meetings.

Finally, another alternative model could be one where students exhibit preferences for equality. Such a model could certainly explain the reduced dispersion in outcomes but it would generally display multiple equilibria, with no predictable impact on average performance. There would be no reason for a generalized fall in performance, unless all students coordinate to the same low-level equilibrium, perhaps because of some peculiar institutional features. Furthermore, preferences for equality would not be enough to explain the observed relationship between the number of meetings and the average and the dispersion of performance. One would have to make further assumptions about how a student's interest for the performance of her friends varies with time spent together. The evidence in Table 11, seems to further reject such a model, as even in ex-ante homogenous pairs we find a reduction in performance, while that model would predict a clustering at a high level or low level, possibly depending on ability.

In general, it is always possible to propose a behavioral model that can explain our empirical findings (in fact, any empirical finding) and one possible reading of our contribution is precisely the ability to explain the patterns of social interactions on the basis of simple economic mechanisms, that are relatively standard and well established in the literature.

6 Conclusions

In this paper we propose a set of models of social interactions that generate testable implications capable of separating them empirically. In particular, we consider three possible mechanisms of interactions: i. a baseline scenario, where peer effects arise because of reference-based utility; ii. a model of joint production and; iii. one of mutual insurance against shocks to time or productivity. While all mechanisms predict a reduction in the dispersion of outcomes within groups (pairs) - what is commonly termed peer effects -, they differ in their implications for average performance, which, compared to the benchmark case, increases under cooperation and decreases with mutual insurance.

We take these simple predictions to the data using information on two cohorts of undergraduate Bocconi students, where we can exploit random variation in the number of meetings between students dictated by repeated random allocation into teaching classes. In this setting, the amount of time any two students spend together in the same lecturing classrooms is exogenous by definition and drives the transition from a decentralized model towards a model of insurance or cooperation. We find that the insurance motive dominates.

Such a result has clear policy implications: in order to increase average performance it would be beneficial to prevent students from sitting in the same class too often. Alternatively, the university could introduce incentives that limit the possibility to engage in mutual insurance and encourage cooperation or joint production.

Although we frame our theoretical discussion as well as the empirical application in the education setting, our analysis is more general and it applies to many different contexts, from teamwork to academic co-authorship to any environment where social interactions could be important. For example, if mutual insurance were proven to be the main mechanism of interaction in team production too, the design of incentive pay schemes, which is one of the fundamental issues in that literature, should take into account the possibility that workers may endogenously react to the introduction of such schemes by engaging in some kind of mutual insurance, thus undoing the expected effect on effort. Similarly, in the production of academic research people explicitly interact through co-authorship, although performance evaluation is typically carried out at the individual level, primarily via tenure decisions but also with the allocation of research funds and awards. This is also a setting that resembles our model very closely.

In the theoretical discussion we have assumed away agents' heterogeneity and endogenous group formation. Although such an assumption is consistent with our testing strategy, where students are randomly allocated to peers (and fixed effects are accounted for), we should emphasize that one important implication of our results is that an individual may choose different types of peers depending on whether the purpose of the group is cooperation or insurance. We also acknowledge that the process of group formation is very important and that the interlink between such process and the choice of the interaction mode should be investigated thoroughly in future research.

References

Akerlof, George A. 1997. "Social distance and social decisions." Econometrica, 65(5): 1005–1027.

Alchian, Armen, and Harold Desmetz. 1972. "Production, Information costs, and Economic Organizations." American Economic Review, 62: 777–795.

Alvarez, Fernando, and Urban J. Jermann. 2000. "Efficiency, Equilibrium, and Asset Pricing with Risk of Default." American Economic Review, 68(4): 775–797.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou. 2006. "Who's who in networks. Wanted: the key player." Econometrica, 74(5): 1403–1417.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2009a. "Social Incentives in the Workplace." Review of Economic Studies, 62: 777–795.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2012. "Team Incentives: Evidence from a Field Experiment." Journal of the European Economic Association. forthcoming.

Bayer, Patrick, Randi Pintoff, and David Pozen. 2009. "Building criminal capital behind bars: Peer effect in juvenile corrections." Quarterly Journal of Economics, 124(1): 105–147.

Bernheim, Douglas B. 1994. "A Theory of Conformity." The Journal of Political Economy, 102(5): 841–877.

Bisin, Alberto, Andrea Moro, and Giorgio Topa. 2011. "The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions." NBER working paper: 17196.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." Journal of Econometrics, 150(1): 41–55.

Brock, William A., and Steven N. Durlauf. 2001. "Interaction-based models." In Handbook of Econometrics. Vol. 5, , ed. James J. Heckman and Edward E. Leamer, Chapter 54, 3297–3380. Elsevier.

Cameron, Colin A., Douglas M. Miller, and Jonah B. Gelbach. 2011. "Robust Inference with Multiway Clustering." Journal of Business and Economic Statistics, 29(2): 238–249.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. 2012. "Inequality at Work: the Effect of Peer Salaries on Job Satisfaction." American Economic review, 102(6): 2981–3003.

Carrell, Scott, Bruce I. Sacerdote, and James E. West. 2012. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." Econometrica. forthcoming.

Coate, Stephen, and Martin Ravallion. 1993. "Reciprocity without commitment : Characterization and performance of informal insurance arrangements." Journal of Development Economics, 40(1): 1–24.

Cochrane, John H. 1991. "A simple test of consumption insurance." Journal of Political Economy, 99(5): 957–976.

Cooley, Jane. 2009. "Desegregation and the achievement gap: Do diverse peers help?" mimeo University of Winsconsin-Madison.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2010. "Identification of Social Interactions through Partially Overlapping Peer Groups." American Economic Journal: Applied Economics, 2(2): 241–275.

De Giorgi, Giacomo, Michele Pellizzari, and William Gui Woolston. 2012. "Class Size and Class Heterogeneity." Journal of the European Economic Association, 10(4): 795–830.

Falk, Armin, and Andrea Ichino. 2006. "Clean evidence on peer effects." Journal of Labor Economics, 24(1): 39–57.

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." The Bell journal of Economics, 13(2): 324–340.

Jackson, Matthew O. 2008. Social and Economic Networks. Princeton University press.

Jackson, Matthew O. 2009. "An Overview of Social Networks and Economic Applications." In The Handbook of Social Economics. , ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson. Elsevier Press. forthcoming.

Jappelli, Tullio, and Luigi Pistaferri. 2006. "Intertemporal choice and consumption mobility." Journal of the European Economic Association, 4(1): 75–115.

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

Kocherlakota, Narayana R. 1996. "Implications of efficient risk sharing without commitment." Review of Economic Studies, 63(4): 595–609.

Lazear, Edward P. 2001. "Educational Production." Quarterly Journal of Economics, 116(3): 777–803.

Ligon, Ethan, Jonathan P. Thomas, and Tim Worrall. 2000. "Mutual Insurance, Individual Savings and Limited Commitment." Review of Economic Dynamics, 3(2): 216–246.

Mace, Barbara J. 1991. "Full insurance in the presence of aggregate uncertainty." Journal of Political Economy, 99(5): 928–956.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." Review of Economic Studies, 60: 531–542.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at work." American Economic Review, 99(1): 112–145.

Mazzocco, Maurizio, and Shiv Saini. 2012. "Testing Efficient Risk Sharing with Heterogenous Risk Preferences." American Economic Review, 102(1): 428–468.

Moffitt, Robert. 2001. "Policy interventions, low-level equilibria and social interactions." In Social Dynamics. , ed. Steven N. Durlauf and Peyton H. Young, Chapter 3. Cambridge: MIT Press.

Moretti, Enrico. 2011. "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales." Review of Economic Studies, 78(1): 356–393.

Moulton, Brent R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." Review of Economics and Statistics, 72(2): 334–338.

Rosen, Sherwin. 1986. "Prizes and Incentives in Elimination Tournaments." The American Economic Review, 76(4): 701–715.

Schulloffer-Wohl, Sam. 2008. "Heterogeneous Risk Preferences and the Welfare Cost of Business Cycles." Review of Economic Dynamics, 11(4): 761–780.

Swamy, P. A. V. B., and S. S. Arora. 1972. "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models." Econometrica, 40(2): pp. 261–275.

Townsend, Robert M. 1994. "Risk and Insurance in Village India." Econometrica, 62: 539–91.

Figures and Tables

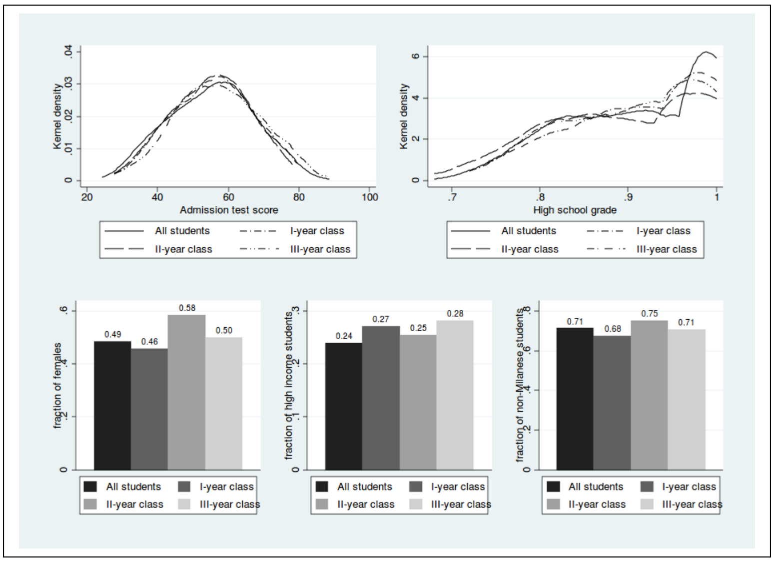

Figure 1: Distribution of selected variables within groups and in the population

Table 1: Compulsory curricula

	Management		Economics		Econ&Finance	
	courses	hours	courses	hours	courses	hours
	[1]	[2]	[3]	[4]	[5]	[6]
Academic year						
Year 1	8	464	8	448	8	464
Year 2	7	384	7	432	7	448
Year 3	8	464	4	256	7	384
Total	23	1,312	19	1,136	22	1,296

Number of courses and hours by program. The detailed list of courses is in the On-line Appendix, Table Web.1.

Table 2: Students' characteristics

Variable	All students N=2,406 mean (std.dev.)	Management N=1,713 mean (std.dev.)	Economics N=224 mean (std.dev.)	Econ&Finance N=496 mean (std.dev.)
1=female	0.415	0.451	0.388	0.299
	-	-	-	
1=original residence outside Milan[a]	0.676	0.664	0.638	0.736
	-	-	-	
1=highest income bracket[b]	0.223	0.235	0.246	0.171
	-	-	-	
Admission test score[c]	65.213	63.702	69.369	68.742
	(15.80)	(15.54)	(16.80)	(15.36)
High school leaving grade[d]	89.542	88.721	89.991	92.324
	(10.19)	(10.30)	(10.89)	(8.89)

[a] Coded to 1 if the student's residence is outside the province of Milan (which is where the university is located).
[b] Coded to 1 if the student's household income falls in the highest bracket for the determination of fees.
[c] Normalized between 0 and 100.
[d] Normalized between 0 and 100 (pass = 60)

Table 3: Characteristics of the student pairs

Variable *(Num. of pairs)*	All *(800,935)*	Management *(732,808)*	Economics *(13,161)*	Econ&Finance *(54,966)*
1=same sex	0.512	0.507	0.521	0.581
1=both males	0.317	0.303	0.372	0.493
1=both females	0.195	0.204	0.149	0.088
1=both from outside Milan[a]	0.449	0.443	0.391	0.538
1=both from Milan[a]	0.113	0.115	0.142	0.072
1=both high income[b]	0.053	0.055	0.056	0.029
1=none high income[b]	0.592	0.586	0.582	0.686
$Abs(test_i - test_j)/test$[c]	0.234	0.236	0.229	0.214
$Abs(HS_i - HS_j)/HS$[c]	0.125	0.127	0.133	0.101
1=ever met	0.375	0.328	0.874	0.873

The table reports the means of the indicated variables.
[a] Coded to 1 if both student's residence is inside (outside) the province of Milan (which is where the university is located).
[b] Coded to 1 if the household incomes of both (none) students falls in the highest bracket for the determination of fees.
[c] The absolute difference in test scores and high school grades between the two students in the pair are normalized by the average test score and high school grade among all students in the same enrollment cohort and degree program.

Table 4: Characteristics of the student pairs (2)

	All pairs			Management			Economics			Econ&Finance		
Variable	year 1	year 2	year 3	year 1	year 2	year 3	year 1	year 2	year 3	year 1	year 2	year 3
Courses in	1.25	2.34	3.52	1.00	1.87	2.86	4.02	7.49	9.48	3.99	7.48	10.96
the same class	(2.91)	(4.02)	(5.07)	(2.64)	(3.51)	(4.39)	(4.00)	(5.32)	(5.68)	(4.00)	(5.31)	(6.37)
Hours in the	72.50	135.00	202.90	57.80	105.50	163.10	225.00	439.40	566.40	231.30	454.70	645.90
same classroom	(168.30)	(232.90)	(293.60)	(153.20)	(198.90)	(250.90)	(224.00)	(311.20)	(336.30)	(232.00)	(322.10)	(375.80)
Mean grade	25.31	25.83	26.57	25.25	25.79	26.55	26.06	25.63	26.27	25.86	26.33	26.85
(weighted)	(1.58)	(1.69)	(1.68)	(1.57)	(1.68)	(1.69)	(1.59)	(1.85)	(1.82)	(1.56)	(1.72)	(1.51)
Coeff. of variation	0.103	0.100	0.094	.103	0.100	0.094	0.095	0.109	0.102	0.096	0.093	0.087
(weighted)	(0.040)	(0.044)	(0.045)	(0.040)	(0.043)	(0.045)	(0.042)	(0.050)	(0.056)	(0.042)	(0.050)	(0.043)
1=grade vectors cross	0.88	0.83	0.84	0.88	0.83	0.84	0.85	0.81	0.68	0.88	0.79	0.84

Mean grade and the coefficient of variation are computed weighting each course by the number of teaching/lecturing hours. Standard errors in parentheses.

Table 5: Distribution of hours together

	Management		Economics		Econ&Finance	
Meetings	%	Hours	%	Hours	%	Hours
Never	67.16	0	12.61	0	12.71	0
Year 1	9.55	464	12.83	448	12.64	464
Year 2	9.51	384	12.34	432	12.48	448
Year 3	9.52	464	12.46	256	12.38	384
Year 1&Year 2	1.37	848	12.6	880	12.37	912
Year 1&Year 3	1.35	928	12.46	704	12.42	848
Year 2&Year 3	1.35	848	12.37	688	12.58	832
Always	0.19	1312	12.33	1136	12.42	1296

Table 6: Time together and academic outcomes

	Mean grade		Coeff. of variation	
	[1]	[2]	[3]	[4]
Hours in the same class	-0.022	-0.027	-0.009	-0.007
multi-clustered z-stat[a]	[-1.46]	[-2.52]	[-1.81]	[-1.18]
network-corrected z-stat[b]	[-8.00]	[-6.38]	[-1.63]	[-0.33]
Student pairs' fixed effects	no	yes	no	yes
Observations	2,402,805	1,601,870	2,402,805	1,601,870
Number of pairs	800,935	800,935	800,935	800,935

[a] The variance-covariance matrix of the estimates is clustered at the level of each student in the pair and the frequency of their meetings.

[b] The variance-covariance matrix of the parameters is estimated element-by-element (using a 10% random sample of students), allowing correlation through the network up to links of distance 4 and further differentiating links across and within the same academic year. See Appendix B for further details.

Both the dependent variables and the explanatory variables are normalized within degree program and academic year cells. All regressions include academic year dummies. The specifications in columns 1 and 3 also include dummies for enrollment cohort and degree program.

Table 7: The crossing property

	1=grade vectors cross	
	[1]	[2]
Hours in the same class	-0.002	-0.004
multi-clustered z-stat[a]	[-0.70]	[-3.15]
network-corrected z-stat[b]	[-2.67]	[-4.00]
Student pairs' fixed effects	no	yes
Observations	2,402,805	1,601,870
Number of pairs	800,935	800,935

[a] The variance-covariance matrix of the estimates is clustered at the level of each student in the pair and the frequency of their meetings.

[b] The variance-covariance matrix of the parameters is estimated element-by-element (using a 10% random sample of students), allowing correlation through the network up to links of distance 4 and further differentiating links across and within the same academic year. See Appendix B for further details.

The explanatory variables are normalized within degree program and academic year cells. All regressions include academic year dummies. The specifications in column 1 also includes dummies for enrollment cohort and degree program.

Table 8: Placebo regressions

	Mean grade	Coeff. of variation	1=crossing
	[1]	[2]	[3]
1=meet anytime in the future	0.001	-0.002	0.001
	[0.29]	[-0.97]	[0.87]
Number of pairs	800,935	800,935	800,935

All outcome variables refer to the first academic year. Both the dependent variables and the explanatory variables are normalized within degree program. All regressions include dummies for enrollment cohort and degree program.

The asymptotic z-statistics in square brackets are two-way clustered at the level of each student in the pair.

Table 9: Robustness check with class effects

	Coeff. of variation		1=grade vectors cross	
	[1]	[2]	[3]	[4]
Hours in the same class	-0.013	-0.014	-0.002	-0.004
	[-2.26]	[-8.87]	[-0.61]	[-4.64]
Student pairs' fixed effects	no	yes	no	yes
Class fixed effects	yes	yes	yes	yes
Observations	2,402,805	1,601,870	2,402,805	1,601,870
Number of pairs	800,935	800,935	800,935	800,935

In columns 1 and 2 both the dependent variables and the explanatory variables are normalized within degree program and academic year cells. In columns 3 and 4 only the explanatory variables are normalized. All regressions include academic year dummies. The specifications in columns 1 and 3 also include dummies for enrollment cohort and degree program.
The asymptotic z-statistics in square brackets are three-way clustered at the level of each student in the pair and the frequency of their meetings.

Table 10: Heterogeneity of the effects across gender compositions

	Mean Grade	Coeff. of variation	Observations
	[1]	[2]	[3]
Gender composition of the pair:			
Same sex	-0.028	-0.008	820,334
	[-2.51]	[-1.22]	
Mixed sex	-0.006	-0.011	781,536
	[-2.50]	[-1.07]	
Both females	-0.014	-0.005	312,150
	[-1.81]	[-0.080]	
Both males	-0.034	-0.014	508,184
	[-2.66]	[-1.56]	

The table reports the fixed-effects estimates of the effect of hours together on the outcome restricting the sample to type of pairs indicated in the first column.
The asymptotic z-statistics in square brackets are three-way clustered at the level of each student in the pair and the frequency of their meetings.

Table 11: Heterogeneity of the effects across the distribution of ability

PANEL A: Mean grade					
		Student i			
	quartiles:	first	second	third	fourth
Student j	first	-0.029	-0.026	-0.028	-0.031
		[-2.27]	[-2.21]	[-2.49]	[-2.35]
		-	-0.025	-0.024	-0.032
	second	-	[-1.96]	[-2.36]	[-2.96]
		-	-	-0.022	-0.024
	third	-	-	[-2.23]	[-2.25]
		-	-	-	-0.033
	fourth	-	-	-	[-2.28]
PANEL B: Coeff. of variation					
		Student i			
	quartiles:	first	second	third	fourth
Student j	first	-0.000	-0.003	-0.009	-0.011
		[-0.01]	[-0.31]	[1.20]	[-1.24]
	second	-	-0.002	-0.008	-0.002
			[-0.19]	[-1.07]	[-0.32]
	third	-	-	-0.012	-0.013
				[-1.13]	[-2.20]
	fourth	-	-	-	-0.007
					[-0.80]

The table reports the fixed-effects estimates of the effect of hours together on the outcome, restricting the sample to pairs of students in the indicated quartiles of the ability distribution. Ability is measured by the entry test score. Quartiles ordered in ascending order.
The asymptotic z-statistics in square brackets are three-way clustered at the level of each student in the pair and the frequency of their meetings.

Appendix A Proofs

Proof of Proposition 1.

Proof. (a) By contradiction, suppose $e_i > e_j$, then $\frac{\partial f(e_i)}{\partial e_i} < \frac{\partial f(e_j)}{\partial e_j}$ and from the first order conditions in equations 10 and 11:

$$f(e_j) - \frac{\gamma}{1-\epsilon} f(e_i) \quad > \quad f(e_i) - \gamma(1-\epsilon) f(e_j) \tag{A1}$$

$$f(e_j)\left[1 + \gamma(1-\epsilon)\right] \quad > \quad f(e_i)\left[1 + \frac{\gamma}{1-\epsilon}\right] \tag{A2}$$

For equations A1 and A2 to be jointly satisfied with $\frac{\partial f(e_i)}{\partial e_i} < \frac{\partial f(e_j)}{\partial e_j}$, it must be that

$$\begin{aligned} 1 + \gamma(1-\epsilon) \quad &> \quad 1 + \frac{\gamma}{1-\epsilon} \\ 1 - \epsilon \quad &> \quad \frac{1}{1-\epsilon} \end{aligned} \tag{A3}$$

which is impossible with $\epsilon < 1$. (b) Subtracting the first order conditions in equations 10 and 11 from each others yields:

$$x_i - x_j = \frac{1}{1+\gamma}\left[\frac{\partial f(e_i)}{\partial e_i} - (1-\epsilon)\frac{\partial f(e_j)}{\partial e_j}\right] > 0 \tag{A4}$$

which is positive, given that $\frac{\partial f(e_i)}{\partial e_i} > \frac{\partial f(e_j)}{\partial e_j}$. ■

Proof for proposition 2.

Proof. The proof is in two steps. First, we show that $\frac{de_i}{d\gamma}$ and $\frac{de_j}{d\gamma}$ must have the same sign. Take the total differential of the first order conditions in equations 10 and 11 and rearrange terms to obtain:

$$\left[\frac{\partial f(e_i)}{\partial e_i} - \frac{\partial^2 f(e_i)}{\partial^2 e_i}\right]\frac{de_i}{d\gamma} \quad = \quad (1-\epsilon) f(e_j) + \gamma(1-\epsilon)\frac{\partial f(e_j)}{\partial e_j}\frac{de_j}{d\gamma} \tag{A5}$$

$$(1-\epsilon)\left[\frac{\partial f(e_j)}{\partial e_j} - \frac{\partial^2 f(e_j)}{\partial^2 e_j}\right]\frac{de_j}{d\gamma} \quad = \quad f(e_i) + \gamma\frac{\partial f(e_i)}{\partial e_i}\frac{de_i}{d\gamma} \tag{A6}$$

which cannot be jointly satisfied if $\frac{de_i}{d\gamma}$ and $\frac{de_j}{d\gamma}$ have different signs. Second, we show that $\frac{de_i}{d\gamma} > 0$ (or alternatively that $\frac{de_j}{d\gamma} > 0$). Combining the two first order conditions in equations 10 and 11 yields:

$$\begin{aligned} \left[\frac{\partial f(e_i)}{\partial e_i} - \frac{\partial^2 f(e_i)}{\partial^2 e_i}\right]\frac{de_i}{d\gamma} \quad = \quad & (1-\epsilon) + \frac{\gamma(1-\epsilon)\frac{\partial f(e_j)}{\partial e_j} f(e_i)}{\frac{\partial f(e_j)}{\partial e_j} - \frac{\partial^2 f(e_j)}{\partial^2 e_j}} + \\ & + \frac{\gamma^2(1-\epsilon)\frac{\partial f(e_j)}{\partial e_j}\frac{\partial f(e_i)}{\partial e_i}}{\frac{\partial f(e_j)}{\partial e_j} - \frac{\partial^2 f(e_j)}{\partial^2 e_j}}\frac{de_i}{d\gamma} \end{aligned} \tag{A7}$$

which shows that $\frac{de_i}{d\gamma} > 0$ if

$$\frac{\partial f(e_i)}{\partial e_i} - \frac{\partial^2 f(e_j)}{\partial^2 e_i} - \gamma^2(1-\epsilon)\frac{\frac{\partial f(e_j)}{\partial e_j}\frac{\partial f(e_i)}{\partial e_i}}{\frac{\partial f(e_j)}{\partial e_j} - \frac{\partial^2 f(e_j)}{\partial^2 e_j}} > 0$$

$$\frac{\partial f(e_i)}{\partial e_i}\left[1 - \gamma^2(1-\epsilon)\frac{\frac{\partial f(e_j)}{\partial e_j}}{\frac{\partial f(e_j)}{\partial e_j} - \frac{\partial^2 f(e_j)}{\partial^2 e_j}}\right] - \frac{\partial^2 f(e_j)}{\partial^2 e_i} > 0 \tag{A8}$$

which is in fact positive. (b) The first result ($\frac{E(x)}{\partial\gamma} \geq 0$) comes immediately from the fact that both e_i and e_j increase with γ. To find the sufficient conditions under which $\frac{CV(x)}{\partial\gamma} \leq 0$, compute the variation in $CV(x)$ as γ changes:

$$\begin{aligned}\frac{dCV(x)}{d\gamma} &= \frac{2}{(x_i+x_j)^2}\left[x_j\frac{dx_i}{d\gamma} - x_i\frac{dx_j}{d\gamma}\right]\\ &= \frac{2(1-\epsilon)}{(x_i+x_j)^2}\left[f(e_j)\frac{\partial f(e_i)}{\partial e_i}\frac{de_i}{d\gamma} - f(e_i)\frac{\partial f(e_j)}{\partial e_j}\frac{de_j}{d\gamma}\right]\end{aligned} \tag{A9}$$

It is easy to show that in equilibrium $f(e_j)\frac{\partial f(e_i)}{\partial e_i} > f(e_i)\frac{\partial f(e_j)}{\partial e_j}$, which implies that the necessary condition to have $\frac{dCV(x)}{d\gamma} < 0$ is $\frac{de_j}{d\gamma} > \frac{de_i}{d\gamma}$ and the higher the difference between $\frac{de_j}{d\gamma}$. and (c) $\frac{de_i}{d\gamma}$ the more likely that $\frac{dCV(x)}{d\gamma} < 0$. In particular, it can be shown that if $\frac{\partial^3 f(\cdot)}{\partial^3 e_a} > 0$ then $\frac{de_j}{d\gamma} > \frac{de_i}{d\gamma}$. In fact, the sufficient condition to have $\frac{dCV(x)}{d\gamma} < 0$ essentially states that $\frac{\partial^3 f(\cdot)}{\partial^3 e_a}$ must be positive and large enough:

$$\frac{\partial^3 f(\cdot)}{\partial^3 e_a} >> 0 \;\Rightarrow\; \frac{de_j}{d\gamma} >> \frac{de_i}{d\gamma} \;\Rightarrow\; \frac{dCV(x)}{d\gamma} < 0 \tag{A10}$$

■

Proof of proposition 3.

Proof. (a) By contradiction, suppose $e_i^C < e_i^D$ and look at how the first order condition in equation 15 changes when effort changes from e_i^C to e_i^D:

$$\begin{aligned}\frac{\partial^2 g(e_i,e_j)}{\partial^2 e_i} &= \frac{\partial g(e_i,e_j)}{\partial e_i}\\ &-\gamma(1-\epsilon)\frac{\partial g(e_j^C,e_i^C)}{\partial e_i^C}(e_i^D - e_i^C)\end{aligned} \tag{A11}$$

If $e_i^C < e_i^D$, the LHS of this equation would be negative and the RHS would be positive, which is impossible. Hence, it must be that $e_i^C > e_i^D$. Similarly for $e_j^C < e_j^D$. ■

(b) given (a) and the complementarity assumption.

Proof of proposition 4.

Proof. Subtracting the first order conditions in equations 15 and 16 to one another, yields:

$$x_i^C - x_j^C = \frac{1}{1+\gamma}\left[\frac{\partial g(e_i^C,e_j^C)}{\partial e_i^C} - (1-\epsilon)\frac{\partial g(e_j^C,e_i^C)}{\partial e_j^C}\right] \tag{A12}$$

Take the total differential of $x_i^C - x_j^C$ from this expression when e_i^D increases to e_i^C and e_j^D increases to e_j^C:

$$\begin{aligned}
d(x_i^C - x_j^C) &= \frac{\partial^2 g(e_i^C, e_j^C)}{\partial^2 e_i^C} de_i + \frac{\partial^2 g(e_i, e_j)}{\partial e_i \partial e_j} de_j - \\
&\quad -(1-\epsilon)\frac{\partial^2 g(e_j^C, e_i^C)}{\partial^2 e_j^C} de_j - (1-\epsilon)\frac{\partial^2 g(e_j, e_i)}{\partial e_j \partial e_i} de_i \\
&= de_i \left[\frac{\partial^2 g(e_i^C, e_j^C)}{\partial^2 e_i^C} - (1-\epsilon)\frac{\partial^2 g(e_j, e_i)}{\partial e_j \partial e_i} \right] - \\
&\quad -de_j \left[\frac{\partial^2 g(e_i, e_j)}{\partial e_i \partial e_j} - (1-\epsilon)\frac{\partial^2 g(e_j^C, e_i^C)}{\partial^2 e_j^C} \right] < 0
\end{aligned}$$

which, given that $de_i > 0$, $de_j > 0$, $\frac{\partial^2 g(e_i,e_j)}{\partial^2 e_i} < 0$ and $\frac{\partial^2 g(e_i,e_j)}{\partial e_i \partial e_j} > 0$, is negative. ■

Proof of proposition 5.

Proof. (a) Simply comparing the first order conditions for the insurance model in equation 19 and the decentralization model in equations 10 or 11 shows that $e^I < e_i^D$. It, then, follows that $E(x^I) < E(x^D) < E(x^C)$. (b) Under full insurance the performance of both students is identical, hence $CV(x^I) = 0$. ■

Appendix B Network-corrected variances

In a network structure all units (nodes) are potentially interlinked, either directly or indirectly. Therefore, when estimating models where the unit of analysis is a network node, like our models in equations 22 and 23, one needs to allow for the potential structure of correlation among the unobservables of such units across the network. Our data are further complicated by two additional features: the pairwise structure and the longitudinal dimension.

As already discussed in the main text, Tables 6 and 7 report our main estimates based on two alternative assumptions about the correlation structure of the error terms in equations 22 and 23. In the multi-clustered case, we allow pairs which have at least one peer student in common to be correlated with each other in an unspecified form and we further cluster the observations at the semi-aggregate level of variation of our main regressor of interest. However, in such specification we abstract from the network structure of the data.

This is taken into account in the second set of z-statistics that are reported in Tables 6 and 7, where we estimate the variance-covariance matrices of the parameters element by element, as in a GLS model. Given the high computational complexity, we perform this exercise on a 10% random sample of our students. Notice that we do not take a 10% random sample of student pairs but we rather go back to the original data, select a 10% random sample of individual students and construct the full network for such a sub-sample, so as to guarantee that the network structure is preserved.

More specifically, in order to compute the network-corrected variance-covariance matrices we first transform the model by applying the standard random-effects or fixed-effects transformations (as we also did for the multi-clustered standard errors), which take into account the longitudinal structure of the data.

Call ν_{ijt} and ϵ_{ijt} the transformed versions of the errors u_{ijt} and e_{ijt} of equations 22 and 23, where we dropped the subscripts c and p for notational ease.[28] Next, for computational simplicity we assume homoskedasticity and we set the variance of both ν_{ijt} and ϵ_{ijt} equal to their standard OLS estimators for all ij and t. Finally, we define the covariance between two generic error terms as follows:

$$Cov(\nu_{ijt}, \nu_{hks}) = \sum_{d=0}^{4}\sum_{l=0}^{2} \rho^{dl} G^{dl}_{(ij)(hk)} \tag{A13}$$

where ρ^{dl} is the part of the covariance due to the presence of one network link of distance d observed either in the same year ($t = s$, $l = 0$), one ($|t - s| = 1$, $l = 1$) or two years apart ($|t - s| = 2$, $l = 2$); $G^{dl}_{(ij)(hk)}$ is the number of such links for the couple of pairs $(ij)(hk)$.

The analog of equation A13 for the fixed-effect model of equation 23 simply replaces ν_{ijt} with ϵ_{ijt} and sets the upper limit of l to 1, since when taking orthogonal deviations one time observation is dropped.[29]

Links of distance one and higher are defined as in the standard network literature: student pairs who are in the same class (in the same year) are nodes of distance one or direct links; pairs who are not directly linked but share one (or more) direct links have links of distance two and so on. Notice that in our framework the network arises due to the repeated random allocations to teaching classes over academic years, hence, within the same academic year, there can only be direct links (distance one). In other words, if two pairs are in the same class in year t they have no links of distance two or higher in the same year, hence $G^{l0}_{(ij)(hk)} = 0$ for $link \geq 2$ and for all $(ij)(hk)$.

In order to take into account the pairwise structure of our data we also define links of distance zero, as those pairs that share either one or the other student.

[28] Notice that those subscripts would be redundant in the derivation of this appendix, given that we maintain the assumption that only the errors of pairs within the same cohort and degree program can be correlated with each other.

[29] For formal correctness the parameters ρ^{dl} should also be labeled differently.

The correlation structure of equation A13 can be easily rationalized with an exogenous process of network formation generated by semi-aggregate shocks that hit multiple students at the same time. Students hit by the same shock are directly connected. The process of class formation in our setting exactly resembles this structure.

To produce the network-corrected variance-covariance matrices we estimate the parameters ρ^{dl} of equation A13 combining observations on the $G^{dl}_{(ij)(hk)}$, which can be computed directly from our data for each couple of pairs $(ij)(hk)$, and OLS estimates of the covariances $Cov(\nu_{ijt}, \nu_{hks})$.

Practically, we simply regress each off-diagonal element of the variance-covariance matrix of the OLS residuals of the (transformed) equations 22 and 23 on the observed $G^{dl}_{(ij)(hk)}$. We use the resulting estimates of the parameters ρ^{dl} to compute the network corrected variance-covariance matrices. Other estimation procedure are possible, such as GMM or minimum distance estimators, but are all equivalent in our setting.[30]

Our sample includes over 800,000 pairs, observed over 3 years, resulting in almost 2.5 million observations. Hence, there are almost 3 trillions unique off-diagonal elements in the variance-covariance matrix of equation 22.[31] Thus, the practical implementation of our procedure to compute network-corrected standard errors is extremely demanding in terms of computing power. For this reason, we select a 10% random subsample of students and consider only the full network for these students when computing the network-corrected estimates.

Since the choice of the mechanism is unlikely to be invariant to the structure of the network, our estimates of the parameters in equations 22 and 23 based on the full or the 10% random sample may not be comparable. For presentational convenience we prefer to report asymptotic z-statistics rather than standard errors, so as to avoid reporting also the coefficients estimated on the random sample, which are, however, available from the authors upon request.

[30]In order to improve efficiency, the entire procedure could also be iterated until some convergence, however, given the computational burden of this exercise, we report the first set of estimates. Moreover, one could further improve the procedure by imposing exogenous constraints on the parameters ρ^{dl}. For example, it would seem reasonable to require $\rho^{dl} > \rho^{\widetilde{d}\widetilde{l}}$, for any $d > \widetilde{d}$ or $l > \widetilde{l}$. Our unconstrained estimates do satisfy such conditions.

[31]When taking orthogonal deviations in equation 23 one time observation for each pair is dropped and the resulting sample is 1/3 smaller.