Understanding Social Interactions:
Evidence from the Classroom*

Giacomo De Giorgi  Michele Pellizzari
Stanford University, NBER and SCID  Bocconi University, IGIER and IZA

First version: November 2008
This: June 18, 2010

Abstract

There is a large literature on social interactions and still little is known about the economic mechanisms leading to the high level of clustering in behavior that is so commonly observed in the data. In this paper we present a model in which agents are allowed to interact according to three distinct mechanisms, and we derive testable implications on the mean and the variance of the outcomes within and across groups. The empirical tests allow us to distinguish which mechanism(s) generates the observed patterns in the data. In our specific application we study the performance of undergraduate students and we find that social interactions take the form of mutual insurance. Such a result bears crucial 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.

JEL Codes: J0, I21
Keywords: Social interactions, peer effects.

*We thank Richard Blundell, Pat Bayer, Matt Harding, Matt Jackson, Enrico Moretti, Eleonora Patacchini. 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 usual disclaimer applies. Corresponding author: Giacomo De Giorgi, Stanford University, 579 Serra Mall, 94305-6072 Stanford, CA. Email: degiorgi@stanford.edu
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) cooperate and exploit complementarities, 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 correlation of outcomes within groups, they have different implications for the average outcome of the production process.

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 cooperation and insurance against it. We show that under cooperation agents exploit complementarities in production and achieve a higher level of output compared to the benchmark. On the contrary, in the mutual insurance scenario the standard moral hazard problem arises and agents exert lower levels of effort because the individual performance is insured against idiosyncratic shocks. As a consequence average output is lower 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 status seeking model (George A. Akerlof 1997). Namely, the utility one enjoys out of a good performance declines if everyone in the group does well. When agents cooperate, the effort levels of all members of the group enter each other’s production functions, thus inducing correlation in outcomes. Finally, the very reason why agents may engage in mutual insurance is the desire to reduce the variability in performance, thus leading to clustering within groups.

We test the implications of our models using data on undergraduate students at Bocconi University, who were randomly assigned to teaching classes. We construct peer groups based exclusively on the random allocation process, thus avoiding issues of endogenous network formation. Moreover, the randomized assignment of students to classes is repeated at the beginning of each academic year, so that the data exhibit a large degree of exogenous variation 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 empirical application. Specifically, we find that the more time the group spends together the lower is the cross-sectional (and longitudinal) variance of the outcomes of its members and the lower is the average performance. Of the three mechanisms considered, insurance is the only one capable of producing both of these results.

Consistently with the literature on risk sharing, these findings are stronger in smaller groups (Joachim De Weerdt & Stefan Dercon 2006, Yann Bramoullé & Rachel Kranon 2007, Attila Ambrus, Markus Mobius & Adam Szeidl 2008, Manuela Angelucci, Giacomo De Giorgi, Marcos A. Rangel & Imran Rasul 2010). The standard problems of information and enforcement make it more difficult to sustaining insurance in larger groups, especially since the contribution of N-th individual in a large insurance group is indeed negligible. Additionally, large groups will be more unstable due to the possibility of forming subgroups or competing coalitions (Garance Genicot & Debraj Ray 2003).

To our knowledge, we are amongst the first to explicitly investigate the economic nature of social interactions. In fact, 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. 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.

We believe that, although we test the implications of our models in the education context, the set-up and topics discussed in this paper are quite general and can be applied to many other areas. For example, in a production team the pressure exerted by the social group, either through sanctions or simple relative utility (Eugene Kandel & Edward P. Lazear 1992), may alleviate the free-riding problem (Armen Alchian & Harold Desmetz 1972, Bengt Holmstrom 1982). Peer pressure also explains the results of Alexandre Mas & Enrico Moretti (2009), where supermarket checkers’ performance improve when they are paired

---

1 A recent wave of studies has approached the identification problem with the use of network structures (Yann Bramoullé, Habiba Djebbari & Bernard Fortin 2009, Antoni Calvó-Armengol, Eleonora Patacchini & Yves Zenou 2009, Giacomo De Giorgi, Michele Pellizzari & Silvia Redaelli 2010).
with high performance checkers. Similarly, peer pressure seems to provide a consistent explanation for
the findings in Armin Falk & Andrea Ichino (2006).

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 principals nor informational asymmetries, how-
ever agents face an uncertain environment characterized by random shocks to their time endowment
(or to their productivity). We believe that many production problems resemble the one studied in this
paper, from manufacturing to construction, to most white collar jobs or coauthoring in academic re-
search, where tenure decisions are made at the individual level but the research activity is often carried
out within groups of co-authors. Similarly, many compensation schemes reward individual performance
even in the presence of interdependencies in the production process.

Without knowledge of the mechanics of social interactions it is often impossible to foresee the effects
of any policy intervention. In this view, the policy implications of our paper are far-reaching. In our
specific 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 introduce incentives to cooperation and limits to the possibility of mutual insurance. In the case of
teamwork the incentives provided by the principal need to account for the possible risk-sharing behavior
among team members (Kandel & Lazear 1992, Bandiera, Barankay & Rasul 2009a, Oriana Bandiera,
Iwan Barankay & Imran Rasul 2009b). For example, in a pool of fruit pickers, as in (Bandiera, Barankay
& Rasul 2009a, Bandiera, Barankay & Rasul 2009b), agents might insure themselves (e.g. against health
shocks or any other random fluctuation in individual productivity) by making sure someone always
shows to work.

The structure of the paper is as follows: Section 2 introduces the key elements of a very simple model
of social interactions; Section 3 describes how to nest the three proposed mechanisms (decentralized
production, cooperation and insurance); in Section 4 we provide a description of the data used in Section
5 to carry out the empirical exercise and take the predictions of the model to the test. Finally, Section 6
concludes.

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.
2 Model

In this section we present a simple model of social interactions, where two ex-ante identical agents \((i, j)\) maximize their utility.

For expositional simplicity 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 generally applied to any setting with social interactions and it is meant to parsimoniously include the key features of the interaction mechanisms that we consider and to produce testable implications that can be taken to the data.\footnote{Possible extensions of the model to \(N\) heterogeneous agents, general utility function (concave), explicit dynamics and more general specifications of uncertainty would complicate the derivation of the testable implications with limited further insights.}

Before moving to the presentation of the model, it is worth emphasizing that the theoretical analysis in this section is crucial for the purpose of uncovering the mechanics of social interactions empirically. It is only by comparing the empirical implications derived from the different models that we can identify the nature of the interactions in the data. A mere empirical analysis without the derivation of a formal model would be incapable of producing any interesting insight into the proposed mechanisms, as the standard regression of individual over group performance would capture a feature that is common to all the economic mechanisms that we consider.

After presenting the building blocks of our theory, we will solve the model under three different scenarios: (i) decentralization or no explicit interactions (Section 2.1); (ii) cooperation or joint production (Section 2.2); (iii) mutual (full) insurance (Section 2.3).

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 \(x_j\), i.e. the utility derived from an A is lower when everyone gets A. An alternative interpretation of this assumption is that students are averse to equality or feel the peers’ pressure. We parameterize this relative utility effect with a loading factor \(\gamma \in [0, 1)\) 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 absence of explicit interactions (provided \(\gamma > 0\)). Eventually, we will work with the following static utility
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:

$$x_i = t_i g(e_i, e_j)$$

While the level of effort is endogenously chosen by the student, time is an exogenous factor subject to idiosyncratic shocks.\(^5\) In this interpretation, it is natural to assume that the two production factors are complements.\(^6\) We assume:

\[\frac{\partial g(e_i, e_j)}{\partial e_i} > 0, \quad \frac{\partial g(e_i, e_j)}{\partial e_j} \geq 0 \quad \text{(production is non-decreasing in efforts);}\]

\[\frac{\partial^2 g(e_i, e_j)}{\partial^2 e_i} < 0, \quad \frac{\partial^2 g(e_i, e_j)}{\partial^2 e_j} \leq 0 \quad \text{(production is non-convex in efforts);}\]

\[\frac{\partial^2 g(e_i, e_j)}{\partial e_i \partial e_j} \geq 0, \quad \text{(non-negative cross-partial derivatives or complementarity of efforts);}\]

\[\frac{\partial g(e_i, e_j)}{\partial e_i} > \frac{\partial g(e_i, e_j)}{\partial e_j}, \quad \text{(the production impact of own effort is larger than that of peers).}\]

When students do not cooperate, $e_j$ does not enter the specification of $x_i$ and, for notational convenience, we define:

$$x_i = t_i g(e_i, 0) = t_i f(e_i)$$

Shocks. Time is equal to a fixed (maximum) endowment normalized to 1 and we allow for a negative shock of size $\epsilon$ 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 $u_i = -\epsilon$ and $u_j = 0$ and with probability $\frac{1}{2}$ we have $u_i = 0$ and $u_j = -\epsilon$. This is equivalent to assuming that the shock to $t$ are perfectly negatively correlated between the two agents. However, the endowment process could be easily generalized and the results

\(^4\)The two agents are homogeneous so $j$ has a corresponding and symmetric utility function. Obviously, $\gamma$ will have to be such that $x_i > \gamma x_j$. The choice of the particular (concave) utility function and the static framework are not crucial for the testable implications produced.

\(^5\)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.

\(^6\)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, given that $t_i$ and $t_j$ are exogenous factors.
would remain qualitatively unchanged as long as there is some degree of idiosyncratic variation in the process. 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 (non-cooperatively) they maximize their utility taking each other’s behavior as given. 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 ($t_j = 1 - \epsilon$):

\[
U_i = \ln \left[ f(e_i) - \gamma (1 - \epsilon) f(e_j) \right] - e_i 
\]

\[
U_j = \ln \left[ (1 - \epsilon) f(e_j) - \gamma f(e_i) \right] - e_j 
\]

Both agents choose their optimal effort level $e^D$ according to the following first order conditions:

\[
\frac{\partial f(e_i^D)}{\partial e_i} = f(e_i^D) - \gamma (1 - \epsilon) f(e_j^D) 
\]

\[
(1 - \epsilon) \frac{\partial f(e_j^D)}{\partial e_j} = (1 - \epsilon) f(e_j^D) - \gamma f(e_i^D) 
\]

It is easy to show that:

**Proposition 1** The optimal effort level is not smaller for the student affected by the negative shock: $e_i^D \leq e_j^D$, with equality holding when $\gamma = 0$.

The proof is in the appendix.

From proposition[1] it is immediate to derive the following:

**Proposition 2** 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 proof can be found in the appendix.
Further, we can show that:

**Proposition 3** \( \frac{de_i}{d\gamma} > 0 \) and \( \frac{de_j}{d\gamma} > 0 \)

Proposition 3 implies that, as the strength of the externalities in academic performance (measured by \( \gamma \)) increases, so does the mean outcome of the group \( E(x) = \frac{1}{2}(x_i + x_j) \).

Understanding the implications of relative utility on the dispersion of the 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:

\[
CV(x) = \frac{SD(x)}{E(x)} = \frac{x_i - x_j}{x_i + x_j}
\]  

(12)

As a first approximation, note that when \( \gamma = 0 \), \( e_i^D = e_j^D \) and \( CV(x^D) = \frac{e_i}{x} > 0 \). As \( \gamma \to 1 \), \( CV(x^D) \to 0 \), with \( x^D = \left( x_i^D, x_j^D \right) \). Thus, the more students care about their relative performance the lower the dispersion in the outcomes. Under some additional assumptions about the production function \( f(\cdot) \), this result generalizes to the entire range of feasible values of \( \gamma \). In the appendix we describe such additional assumptions in detail.

The following proposition describes the effect of relative utility evaluations on the mean and the dispersion of outcomes:

**Proposition 4**

- \( \frac{\partial E(x^D)}{\partial \gamma} \geq 0 \) for any admissible \( \gamma \in [0, 1) \);
- under the regularity conditions set out in the appendix, \( \frac{\partial CV(x^D)}{\partial \gamma} \leq 0 \) for any admissible \( \gamma \in [0, 1) \).

Proposition 4 shows that our simple model produces what is commonly termed peer effects also in the absence of any active interaction, simply by virtue of relative utility. We will maintain this case as the baseline scenario.

### 2.2 Cooperation or joint production

In this section we modify the model under the assumption that students cooperate with each other in the production process. 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)\(^8\).

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

\[
U_i = \ln [g(e_i, e_j) - \gamma (1 - \epsilon) g(e_j, e_i)] - e_i \quad (13)
\]

\[
U_j = \ln [(1 - \epsilon) g(e_j, e_i) - \gamma g(e_i, e_j)] - e_j \quad (14)
\]

Define \(e^C\) the optimal effort under cooperation and compute the first order conditions for the maximization of equations 13 and 14:

\[
\frac{\partial g(e^C_i, e^C_j)}{\partial e_i} = g(e^C_i, e^C_j) - \gamma (1 - \epsilon) g(e^C_j, e^C_i) \quad (15)
\]

\[
(1 - \epsilon) \frac{\partial g(e^C_j, e^C_i)}{\partial e_j} = (1 - \epsilon) g(e^C_j, e^C_i) - \gamma g(e^C_i, e^C_j) \quad (16)
\]

Using equations 15 and 16, it is relatively easy to derive the following proposition, where we define \(e^D\) the optimal effort level under decentralization:

**Proposition 5**

- \(e^C_a > e^D_a\) for any \(a = \{i, j\}\);

- **Average performance under cooperation is larger than in decentralization**: \(E(x^C) > E(x^D)\), where \(x^C = \left[ g(e^C_i, e^C_j), (1 - \epsilon) g(e^C_j, e^C_i) \right]\) and \(x^D = \left[ f(e^D_i), (1 - \epsilon) f(e^D_j) \right]\) are the equilibrium outcomes under cooperation and decentralization, respectively.

The proof, in the appendix, is trivial, given that the level of effort is larger under cooperation than decentralization for both students. Another important result of this model is described in the next proposition:

**Proposition 6** \((x^C_i - x^C_j) < (x^D_i - x^D_j)\), hence \(SD(x^C) < SD(x^D)\) as well as \(CV(x^C) < CV(x^D)\).

\(^8\)Edward P. Lazear & Kathryn Shaw (2007) document the importance of teamwork in the US economy. In the same vein, Stefan Wuchty, Benjamin F. Jones & Brian Uzzi (2007) provide evidence on the ever growing importance of teamwork in academic production.
The formal proof of proposition 6 is in the appendix.

In sum, this section shows that, under cooperation, students’ outcomes are higher on average and less dispersed than under decentralization.\footnote{The same result can be derived for a cooperative model in which the agents maximize the sum of their individual utilities (or a linear combination, if they were heterogeneous). In such a model, the presence of production complementarities still leads to higher effort levels (compared to decentralization), likewise in our cooperation model. Additionally, the concavity of the production function induces a larger increase in effort for agent \( j \) than agent \( i \), thus leading to a reduction in the dispersion of outcomes.}

2.3 Insurance

The third and last mechanism of interaction that we consider is mutual insurance (John H. Cochrane 1991, Barbara J. Mace 1991, Robert M. Townsend 1994) against idiosyncratic shocks to time or productivity, e.g. health shocks or random distractions. 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 endowment (or productivity). Failing an exam would imply delaying one’s graduation as well as a fall in GPA. Hence, one would be better off if she could reduce the possibility of a negative outcome by exchanging notes or explanations of class material with her classmates.

As already mentioned, we choose an extremely simple structure of the shocks but the implications of the model are robust to alternative structures. 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 endowment. Moreover, given that our agents are ex-ante identical, we set identical Pareto weights. 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 3. Notice, however, that even under limited commitment, students can achieve full insurance for some parametrization of the primitives of the model, for example when the punishment for deviating from a full insurance arrangement is large enough. Ultimately, the testable implications of our model would remain qualitatively unchanged also with partial insurance (Richard Blundell, Ian Preston & Luigi Pistaferri 2008).

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 optimal level of effort is the same for both of them: $e_i^I = e_j^I = e_I^I$. Formally,

$$U_i = U_j = U = \ln \left[ tg(e, e) - \gamma tg(e, e) \right] - c$$

$$t = (1 - \frac{e}{2})$$

The corresponding first order condition is:

$$\frac{\partial f(e_I^I)}{\partial e_I} = f(e_I^I)$$

and equilibrium performance is:

$$x_I^I = (1 - \frac{e}{2}) f(e_I^I).$$

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 time and states. This observation has one important 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). In Section 5, we take also this prediction to the empirical test.

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

**Proposition 7**

- $e_a^I < e_a^D < e_a^C$ for any $a = \{i, j\}$, which immediately implies $E(x^C) > E(x^D) > E(x_I^I)$;
- $CV(x^I) < CV(x^C) < CV(x^D)$.

The proof is fairly straightforward and can, once again, be found in the appendix. The intuition is that under full insurance there is a clear moral hazard problem which pushes down the level of efforts while decreasing the cross-sectional variance to the lowest possible level as the idiosyncratic shocks.

---

10 Such prediction would be however incorrect if the agents had heterogenous risk preferences (Sam Schulloffer-Wohl 2008, Maurizio Mazzocco & Shiv Saini 2009).
are wiped out. As we show in the Appendix 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.

3 Transition

In this section we briefly discuss how the three mechanisms that we considered in our model in Section 2 can be nested. Consistently with our empirical application where students initially do not know each other and are randomly allocated to teaching classes, we assume that decentralization is the baseline scenario and transition to a model of cooperation and/or insurance is triggered by repeated interactions in the classroom. Although such a transition process could be modeled more structurally, including a full dynamic optimization, we take a reduced form approach that allows us to focus on the main intuitions for our empirical exercise.

First, consider cooperation and assume that the ability to cooperate 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 cost that assumes 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 (Coate & Ravallion 1993) that describes the utility loss derived from deviating from the insurance scheme. Under such an interpretation, the size of the punishment increases with the number of random 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 have an incentive to renge 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 Nash equilibrium would be to renge 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 $m$.

The relative shape of the functions $c(m)$ and $h(m)$ defines which of the models, cooperation or insurance, arises first. We do not take a particular stand on this particular issue as the type of mechanism at work is ultimately an empirical matter.

In sum, our testable predictions imply the following:

\[
\begin{align*}
\text{If } & \frac{\partial \text{Var}(x)}{\partial m} < 0 \text{ and } \frac{\partial \text{E}(x)}{\partial m} > 0 \Rightarrow \text{Cooperation is the prevailing mechanism;} \\
\text{If } & \frac{\partial \text{Var}(x)}{\partial m} < 0 \text{ and } \frac{\partial \text{E}(x)}{\partial m} < 0 \Rightarrow \text{Insurance is the prevailing mechanism.}
\end{align*}
\]

The reduction in the (within-group) dispersion of performances as a function of the time spent together indicates the existence of peer effects, a feature that is common to all models, while the relation between number of meetings and average performance in the group allows to distinguish between cooperation and insurance.

4 The data

In this section we describe the data we use in Section 5 to test the theoretical predictions of our models.

The data come from the administrative archives of Bocconi University, an institution of higher education offering majors in Economics and Management (De Giorgi, Pellizzari & Redaelli 2010, Giacomo De Giorgi, Michele Pellizzari & William Gui Woolston 2009). The important feature of these data for the purpose of our application is that students were (repeatedly) randomly assigned to teaching classes for their compulsory courses.

More specifically, we will focus on the cohort of students who first enrolled at Bocconi in the academic year 2000-2001. 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 degree. For the earlier cohorts we do not have information on the class identifiers, which are essential to construct peer groups. Moreover, Bocconi reformed the structure of its programs twice during the period covered by these data, first in 1999-2000 and then 2001-2002. Hence, to avoid comparing cohorts across different systems, we consider only the 2000-2001 cohort.

At the time covered by our data, Bocconi offered 7 possible majors, with only 3 of them, those
used in the analysis, enrolling a large enough number of students to require multiple teaching classes: *Economics, Management, Economics and Finance*. 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, excluded from the analysis, following some program-specific guidelines.

The complete academic structure of the three degree programs that we consider is described in Table B.1 in Appendix B. The table reports the list of all compulsory courses for each of the three programs (majors), divided by academic year and broad subject areas. Importantly, the table also shows the number of lecturing hours for each course. On average a student should attend more than 450 hours of lectures in each academic year, so that students who attended all the courses together would spend a considerable amount of time sitting in the same classroom.

The allocation of students to teaching classes was based on a completely random computer algorithm and was repeated at the beginning of each academic year, when each student was informed of her class number and was instructed to take lectures for all courses taught during the year in the class corresponding to her assigned number. 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, finally, the same procedure applied to the compulsory courses of the third year.

Since the types and the sequence of compulsory courses differ across degree programs, the allocation to class was conditional on major, so that Economics students would take classes only with Economics students and similarly for the other majors. Notice also that, in our empirical analysis we always control for major fixed effects so that we never use variation generated by the students endogenous choice of degree program. Effectively we exploit exclusively variation generated by the random allocation mechanism.

Elective courses, which we do not consider in our analysis, were usually smaller in size and could easily be taught in a single class. Another important reason, aside from natural self-selection, why we do not consider elective courses is that compulsory exams were graded centrally by a set of graders and

---

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

12 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.
not by the specific teacher. Moreover, the content of the course was also harmonized across the classes of
the same degree program and teachers were all required to adopt the same textbook and follow the same
syllabus. Very little room was left to the individual teacher to adapt the course to his/her preferences.\textsuperscript{13}

Students were allocated into several classes for the explicit purpose of keeping class size to a mini-

mum 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.

Our definition of peers is based on classmates. Specifically, we define the peer group of a student \( i \) as

those students who are randomly allocated to the same teaching class as \( i \). Hence, \( i \)'s peer group varies

over academic years according to the class to which he/she is allocated. Moreover, when constructing
group variables (e.g. average grade, grade dispersion, and so on) we weight each peer by the number of

hours he/she spent in the same classroom as \( i \), either in the current or in the past academic years. This
generates group average and group dispersion of grades that vary also across peers of the same group.\textsuperscript{14}

Consider a simple example with just three students - A, B and C - who are assigned to the same class

in the second year and, hence, are peers to each other. Further suppose that, while A and B were also

assigned to the same class in the first year, student C took all the courses of the first year in a different

class. As a consequence, B is given a higher weight than C in the construction of the group mean and
dispersion of grades for student A.

More formally, we calculate group attributes as follows:

\[
x_{it}^{G_i} = E(x \mid t, G_{it}) = \frac{\sum_{j \in G_{it}} m_{ijt} x_{jt}}{\sum_{j \in G_{it}} m_{ijt}} \tag{21}
\]

where we indicate with \( x_{it} \) a generic variable (say, GPA) for student \( i \) in year \( t \) and with \( x_{it}^{G_i} \) being
the corresponding group mean of the same variable within the group \( G_{it} \) of \( i \)'s peers in year \( t \), i.e. the
students who are assigned to the same class as \( i \) in year \( t \); \( m_{ijt} \) is the number of hours students \( i \) and \( j \)

\textsuperscript{13} There still was some variation in course content across degree programs. For example, the mathematics course taught to
students in the Economics program was slightly more advanced than the others. Nevertheless, all students in the Economics
program were taught the same course, regardless of their assigned class.

\textsuperscript{14} The only exception are observations in the first academic year, when all those who sit in the same class have identical
mean and dispersion of grades.
have spent in the same class since their enrolment at Bocconi and until year \( t \) (included).\(^{15}\)

Applying the definition \(^{21}\) to our previous simple example, the groups of students A, B, and C in year one are all identical - i.e. \( G_{A1} = G_{B1} = G_{C1} \) - while in year two \( G_{A2} = G_{B2} \neq G_{C2} \), so that \( m_{AB2} > m_{AC2} = m_{BC2} \) although \( m_{AB1} = m_{AC1} = m_{BC1} \). Consequently, despite their being in the same group in year two, \( x_{G_{A2}} \neq x_{G_{C2}} \). With three repeated random allocations and with the number of teaching hours that varies across academic years both within and across degree programs, our data are such that the weighting scheme generates peer attributes that vary essentially at the level of the single student, with the only exception of academic year 1.\(^{18}\)

Given the importance of the class allocation for our empirical analysis, we start presenting some evidence consistent with its being random.

[FIGURE 1 HERE]

In Figure 1 we compare the distribution of some selected observable exogenous characteristics in the entire cohort and within the groups of peers of a randomly selected student in each of the three initial academic years\(^ {17}\). As it is evident from the graphs, the distributions look very similar, a visual impression that is confirmed by formal testing\(^ {18}\).

Overall, there are 12 classes per academic year: 8 classes in Management, 2 in Economics and 2 in Economics and Finance. As a byproduct of this particular repeated randomization some students met more often than others as well as in groups of different sizes (De Giorgi, Pellizzari & Woolston 2009). Table 1 shows some descriptive statistics on the groups. Overall, the average class size is slightly above 100 students and it does not change substantially over the three academic years. Since students are very unevenly distributed across degree programs, class size does vary across them with much smaller groups in Economics and larger in Management and Management and Finance\(^ {19}\).

Different programs are also characterized by different probabilities of being assigned to the same class repeatedly over the years. For example, in Economics, where there are only 85 students divided into two classes, the average number of years spent in the same class is 1.68 in the second year and

---

\(^{15}\)By definition, \( m_{ij} = m_{ji} \)

\(^{16}\)Notice that in this simple example we have that \( x_{G_{A2}} = x_{G_{B2}} \), although our actual data show enough variation to exclude that any two peers have exactly the same group variables (after year one).

\(^{17}\)Similar tests for the same data (although using partially different cohorts) are presented in De Giorgi, Pellizzari & Redaelli (2010) and De Giorgi, Pellizzari & Woolston (2009).

\(^{18}\)We perform Kolmogorov-Smirnov tests for the comparison of the distributions of the admission test scores and high school grades and tests for the equality of proportions for all the other variables.

\(^{19}\)When computing class size we do not weight peers by the time spent together.
2.25 in the third. These figures are comparable to the 1.66 and 2.25 for students in Economics and Finance, who are also divided into two classes, but decrease substantially for Management, where the 848 students are allocated into 8 random classes.

Two students who are assigned the same class identifier by the random algorithm spend a substantial number of hours sitting in the same classroom. As already mentioned, the courses of a single academic year account for around 450 hours. Over the years, however, since students may be allocated repeatedly to the same class, the number of hours that two random classmates have spent in the same room increases to around 550 in the second and almost 800 in the third year. Moreover, given the differences across degree programs in the likelihood of repeated meetings, in Economics and Economics and Finance these figures go up to around 750 in the second year and to over 1,000 in the third.

Table 2 reports a series of descriptive statistics for the observable characteristics of the students, broken down by degree program. Our data contain very detailed information, including the students’ performance in each exam, the date in which each exam was sat, demographic information, family income and, importantly, the result of a cognitive ability test that every student takes as part of the admission process. On average 45% of students are females, and mostly come from outside the province of Milan. Overall, students in the different degrees look on average rather similar, apart from a larger share of females in the management program as well as lower average entry test score and household income for this particular program.

Finally, as our analysis in Section 5 focuses mainly on the evolution of both the average and the dispersion of academic outcomes within groups and across academic years, Table 3 shows the basic characteristics of these variables. Namely, we report in the table the mean (std. dev.) and the coefficient of variation (std. dev.) of the exam grades within the group of peers of each student, broken down by academic year and degree program. Consistently with the random nature of the groups, 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.[20]

[TABLE 1 HERE]

[TABLE 2 HERE]

[20] 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.
5 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 track. Following our theoretical discussion in Section 2, different mechanisms of social interactions have different implications on such evolution as summarized at the bottom of Section 3.

In sum, our testable predictions imply the following:

If \( \frac{\partial \text{Var}(x)}{\partial m} = 0 \) and \( \frac{\partial \text{E}(x)}{\partial m} = 0 \) \( \Rightarrow \) Decentralization is the prevailing mechanism;

If \( \frac{\partial \text{Var}(x)}{\partial m} < 0 \) and \( \frac{\partial \text{E}(x)}{\partial m} > 0 \) \( \Rightarrow \) Cooperation is the prevailing mechanism;

If \( \frac{\partial \text{Var}(x)}{\partial m} < 0 \) and \( \frac{\partial \text{E}(x)}{\partial m} < 0 \) \( \Rightarrow \) Insurance is the prevailing mechanism.

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 number of meetings and average performance in the group allows to distinguish between cooperation and insurance. where \( m \) is a measure of the number of meetings or hours students in the group spend together. Namely, as interactions become stronger, we expect the average grade to increase if cooperation 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 coefficient of variation of grades within groups to decrease as the number of meeting or hours spent together increases when the prevailing mechanism is either cooperation 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.\(^\text{21}\)

We test these predictions running the following regression (results in Table 4 and Figure 2):

\[
y_{it} = \alpha_0 + \alpha_1 m_{it} + \alpha_2 s_{it} + \alpha_3 [m_{it} \times s_{it}] + \zeta_{tc} + v_i + u_{it} \tag{22}\]

(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 these 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.

\(^\text{21}\)If we allow \( \gamma \) to be increasing in \( m \), i.e. the relative importance of peers’ performance increases in the number of meetings, we have that \( \frac{\partial \text{E}(x)}{\partial m} > 0 \) and \( \frac{\partial \text{CV}(x)}{\partial m} < 0 \) as shown in Proposition 4.
where \( y_{it} \) is, alternatively, the (log of the) average or the dispersion of performance (GPA) of the group of student \( i \) in each of the first three academic years \( t \); \( m_{it} \) measures the number of meetings or (log of) hours student \( i \) has spent on average with members of her own group; \( s_{it} \) is the (log) size of the individual peer group. Additionally, we control for a full set of interacted time by major effects, \( \zeta_{tc} \). \( v_i \) is a student specific effect, that we are going to treat either as a fixed (columns 1 and 3) or a random term (columns 2 and 4), and \( u_{it} \) is the usual error. The descriptive statistics of our dependent variables are those reported in Table 3.\(^{22}\)

Consistently with the idea that coordination and monitoring problems are more difficult to solve in larger groups (De Weerdt & Dercon 2006, Bramoullé & Kranon 2007, Ambrus, Mobius & Szeidl 2008, Angelucci et al. 2010), in this specification we also allow the effect of \( m_{it} \) to vary with the size of one’s peer group. The inclusion of the size of one’s group in the set of control variables is important also for identification purposes. In fact, the meeting probability of any two students increases in the size of the class and, given the negative effects of class-size on performance estimated in De Giorgi, Pellizzari & Woolston (2009), controlling for \( s_{it} \) avoids any possible bias due to spurious correlation.\(^{23}\)

Otherwise, the identification of the relevant parameters in equation\(^{22}\) is straightforward. Given that the right hand side variables of interest are exogenous (by construction, i.e. the random allocation) we need not worry about possible treats to the identification of the parameters of interest, i.e. \( \alpha_1, \alpha_2 \) and \( \alpha_3 \). Consistently with this interpretation, we obtain very similar results when we estimate equation\(^{22}\) considering the student specific term \( v_i \) a fixed- or a random-effect.

Once again, in estimating the relation between individual performance and the number of hours spent with one’s peers, we expect \( \hat{\alpha}_1 + \hat{\alpha}_3 s_{it} < 0 \) if insurance is the prevailing mechanism. If, instead, we estimate \( \hat{\alpha}_1 + \hat{\alpha}_3 s_{it} > 0 \), then we would conclude that cooperation is the economic driver of classroom interactions. It can also be that both mechanisms operate at the same time so that we are really only able to test which one dominates.

It could well be that some students cooperate, others insure while others act in isolation, an equilibrium that we do not discuss in our theoretical analysis where, for simplicity, we maintain the assumption of homogenous agents. Nevertheless, even in a model with heterogeneous agents who select themselves

\(^{22}\)As an alternative to considering the average peer grade as a dependent variable, we could also test our first empirical implication looking at the individual average grade in each academic year. We prefer the current specification because it is comparable to the approach we use to compute grade dispersion within groups, to which there are no obvious alternatives.

\(^{23}\)Further, the results in Table 4 suggest that the negative relation between the number of meetings and the average performance are stronger in small rather than in large groups, the opposite of what the negative effect of class size would produce.
into different equilibria, we would still have that a negative relation between the mean GPA $y_{it}$ and the number of meetings $m_{it}$ indicates that insurance prevails as interaction becomes more intense.

[TABLE 4 and FIGURE 2 HERE]

The results in Table 4 clearly show that insurance dominates in smaller groups. While the negative relationship between dispersion in performance and the (random) number of hours students in a group spend together would be consistent with both insurance and cooperation (or a combination of the two), the only mechanism that can rationalize the decline in average group performance is insurance. Moreover, the magnitude of the effects is non trivial: for a group of small size (in the bottom 10%, corresponding to approximately 42 students) the elasticity is -0.026 and statistically significant. Such an elasticity remains negative up to group sizes of almost 90 students (just around the 20th percentile of the distribution of class size) and negative and significant (at the 90% level) up to groups of approximately 55 students, i.e. around the 15th percentile of the distribution of class size, when it is equal to -0.017. For larger classes the effect becomes positive but indistinguishable from zero. Figure 2 reports the estimated elasticities (based on the fixed-effect results reported in column 1) and the corresponding 95% and 90% confidence intervals for all levels of (log) group size that are observed in our data.

Looking more specifically at the cross-sectional dispersion in GPA within groups, as measured by the coefficient of variation, we find that the dispersion in grades falls by 1% for an increase of 10% in the hours spent together for groups of 55 students, which is in the bottom 20% of the distribution of sizes; while the same change has no effects on dispersion for average sized groups.

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 grades of students in the same group 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 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

---

24These computations are based on the fixed-effect specification of Table 4, column 3.
25As we already mentioned, this proposition would be incorrect with heterogenous risk preferences. In such a situation, the crossing property would over-reject the null of full insurance even when true (Mazzocco & Saini 2009).

In our specific application, we expect students to insure more and more as they spend more and more time together, eventually approaching full insurance (see Section 3). We test this prediction by constructing pairs or dyads of students and estimating the following equation:

$$cross_{ij,t} = \beta_0 + \beta_1 m_{ij,t} + \tau_t + \xi_c + \eta_{ij} + e_{ij,t}$$ (23)

where $cross_{ij,t}$ is a 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 or before year $t$; $m_{ij,t}$ measures the number of hours $i$ and $j$ have spent in the same class from the time on enrollment until year $t$; $\tau_t$ is a year effect; $\xi_c$ is a program fixed effect; $\eta_{ij}$ is a pair fixed effect and $e_{ij,t}$ is a random error term.

The test above is valid even if insurance takes place at the group level and not necessarily among pairs of students. In fact, for the crossing property to be valid it does not matter whether insurance takes place at the group or at the pairwise level. Consider the simple three-students (A, B, and C) example and assume that the pairs A-B and B-C engage in bilateral full insurance while the pair A-C does not. In such a situation, the ratio of the outcomes of A and B would be always (across time and states of the world) equal to the ratio of the Pareto weights or, without loss of generality, equal to 1 in our setting with identical weights, the same reasoning applies to the (ratio of) outcomes of B and C. Hence, again assuming identical Pareto weights (without loss of generality), also A and C would experience the same outcomes. In other words, bilateral insurance implies group insurance (Joachim De Weerdt & Marcel Fafchamps 2010).

We estimate equation (23) with a linear probability model, both considering $\eta_{ij}$ as a fixed or a random effect, and we adjust the standard errors for clustering at the student pair level. The results are reported in Table 5, where we also experiment with a quadratic function of $m_{ij,t}$. Figure 3 draws the predicted quadratic relationship of $cross_{ij,t}$ and $m_{ij,t}$ using the fixed-effect estimates of column 2 in Table 5.

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. Such relationship also appear

---

26Results are significant, although only at the 90% level, also when adopting the two-way clustering approach of A. Colin Cameron, Jonah B. Gelbach & Douglas L. Miller (2004) which is however known to have poor properties in small samples.
to be nonlinear. For a better assessment of the magnitude of these effects, the graphs in Figure 3 plot the derivative of $\text{cross}_{ij,t}$ with respect to $m_{ij,t}$, for the entire range of observed values of $m_{ij,t}$.

The left panel of Figure 3 indicates that such derivative is positive but never significantly different from zero until about 250 hours of common lectures, which is the minimum among pairs that do meet at least once. Then, the effect becomes negative and statistically significant and it ranges from 0 to -0.02 percentage points over an average of approximately 11%.

The right panel of Figure 3 reports exactly the same results as the left panel with the only difference that the variable on the x-axis has been normalized, so that the effects can be interpreted in terms of standard deviation changes. The graph shows that the effect of hours spent together on the probability of crossing is negative and significantly different from zero roughly within 0 to 3 standard deviations above the mean.

For consistency, we also run the equivalent of equation 22 on the set of student pairs. The results are largely consistent with the previous findings and are reported in Table B2 in Appendix B.

5.1 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 possible that other mechanisms are consistent with our empirical evidence. 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 would be hard pressed to explain the drop in the cross-sectional variance, as each student would be subject to her own shock. 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.

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 4 show exactly the opposite. 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.

Finally, another model able to explain the observed data would 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 students clustering around either high or low performance. 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.

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 possibility to explain the pattern of social interactions on the basis of simple economic mechanisms that are relatively well established in the literature.

6 Conclusions

In this paper we propose a set of models of social interactions able to generate testable implications, which can be used to separate 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 cooperation 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 - 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 bring these simple predictions to the data using information on a cohort 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, in particular in smaller groups.

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.

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 team work to academic co-authorship to any environment where social interactions are 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 biggest 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 how one of the important result of our paper is that an individual may choose different types of peers depending on whether the purpose of the group is cooperation or insurance. We also recognize that the processes of group formation (Coralio Ballester, Antoni Calvó-Armengol & Yves Zenou 2006, Genicot & Ray 2003) are very important and that our simplifying assumption should be relaxed and investigated thoroughly in future research.
References


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

Figure 2: Interaction effects of hours and group size

These effects are based on the results reported in Table 4, columns 1 (for the left panel) and 3 (for the right panel).
These effects are based on the results reported in Table 5, column 2.

Table 1: Characteristics of the groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>first year</th>
<th></th>
<th>second year</th>
<th></th>
<th>third year</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>All students (n=1,153)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>group size</td>
<td>101.88</td>
<td>18.61</td>
<td>101.36</td>
<td>17.81</td>
<td>101.25</td>
<td>17.44</td>
</tr>
<tr>
<td># of years in the same class</td>
<td>1.00</td>
<td>0.00</td>
<td>1.34</td>
<td>0.20</td>
<td>1.64</td>
<td>0.37</td>
</tr>
<tr>
<td># of hours in the same class</td>
<td>462.82</td>
<td>4.18</td>
<td>556.36</td>
<td>119.23</td>
<td>773.12</td>
<td>196.68</td>
</tr>
<tr>
<td>Management (n=848)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>group size</td>
<td>105.90</td>
<td>9.31</td>
<td>105.33</td>
<td>6.13</td>
<td>105.22</td>
<td>4.78</td>
</tr>
<tr>
<td># of years in the same class</td>
<td>1.00</td>
<td>0.00</td>
<td>1.22</td>
<td>0.05</td>
<td>1.42</td>
<td>0.07</td>
</tr>
<tr>
<td># of hours in the same class</td>
<td>464.00</td>
<td>0.00</td>
<td>486.62</td>
<td>24.17</td>
<td>660.72</td>
<td>39.42</td>
</tr>
<tr>
<td>Economics (n=85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>group size</td>
<td>43.20</td>
<td>8.38</td>
<td>41.65</td>
<td>2.51</td>
<td>41.51</td>
<td>0.50</td>
</tr>
<tr>
<td># of years in the same class</td>
<td>1.00</td>
<td>0.00</td>
<td>1.68</td>
<td>0.12</td>
<td>2.25</td>
<td>0.11</td>
</tr>
<tr>
<td># of hours in the same class</td>
<td>448.00</td>
<td>0.00</td>
<td>736.31</td>
<td>54.76</td>
<td>979.80</td>
<td>91.39</td>
</tr>
<tr>
<td>Economics and Finance (n=220)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: The crossing property
<table>
<thead>
<tr>
<th>group size</th>
<th>109.08</th>
<th>3.01</th>
<th>109.15</th>
<th>4.01</th>
<th>109.04</th>
<th>2.00</th>
</tr>
</thead>
<tbody>
<tr>
<td># of years in the same class</td>
<td>1.00</td>
<td>0.00</td>
<td>1.66</td>
<td>0.02</td>
<td>2.25</td>
<td>0.08</td>
</tr>
<tr>
<td># of hours in the same class</td>
<td>464.00</td>
<td>0.00</td>
<td>755.65</td>
<td>8.24</td>
<td>1126.51</td>
<td>57.46</td>
</tr>
<tr>
<td>Variable</td>
<td>mean</td>
<td>s.d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>--------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All students</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=female</td>
<td>0.45</td>
<td>(0.50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=original residence outside Milan (province)</td>
<td>0.73</td>
<td>(0.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross annual household income   (^1)</td>
<td>32,561</td>
<td>(27,369)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=highest income bracket   (^1)</td>
<td>0.23</td>
<td>(0.42)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admission test score   (^2)</td>
<td>56.26</td>
<td>(12.77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school leaving grade  (^3)</td>
<td>0.92</td>
<td>(0.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA (compulsory courses only) (^4)</td>
<td>26.37</td>
<td>(2.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Management</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=female</td>
<td>0.49</td>
<td>(0.50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=original residence outside Milan (province)</td>
<td>0.71</td>
<td>(0.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross annual household income   (^1)</td>
<td>31,578</td>
<td>(27,355)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=highest income bracket   (^1)</td>
<td>0.24</td>
<td>(0.43)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admission test score   (^2)</td>
<td>55.26</td>
<td>(12.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school leaving grade  (^3)</td>
<td>0.91</td>
<td>(0.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA (compulsory courses only) (^4)</td>
<td>26.30</td>
<td>(2.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Economics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=female</td>
<td>0.39</td>
<td>(0.49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=original residence outside Milan (province)</td>
<td>0.72</td>
<td>(0.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross annual household income   (^1)</td>
<td>35,171</td>
<td>(30,077)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=highest income bracket   (^1)</td>
<td>0.29</td>
<td>(0.46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admission test score   (^2)</td>
<td>57.66</td>
<td>(12.60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school leaving grade  (^3)</td>
<td>0.91</td>
<td>(0.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA (compulsory courses only) (^4)</td>
<td>26.41</td>
<td>(2.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Economics and Finance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=female</td>
<td>0.32</td>
<td>(0.47)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=original residence outside Milan (province)</td>
<td>0.78</td>
<td>(0.41)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross annual household income   (^1)</td>
<td>35,340</td>
<td>(26,080)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=highest income bracket   (^1)</td>
<td>0.16</td>
<td>(0.37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admission test score   (^2)</td>
<td>59.56</td>
<td>(13.50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school leaving grade  (^3)</td>
<td>0.94</td>
<td>(0.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA (compulsory courses only) (^4)</td>
<td>26.63</td>
<td>(1.98)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. If a student declares that household income falls in the highest income bracket no further information is collected. The reported mean and standard deviation of gross annual income are computed using only the available information. The figures are in current Euros.
2. Normalized between 0 and 100.
3. Normalized between 0 and 100 (pass = 60)
4. Exams are graded on a scale 0 to 30, pass = 18.
### Table 3: Mean and dispersion of academic outcomes within groups

<table>
<thead>
<tr>
<th></th>
<th>All students</th>
<th>Management</th>
<th>Economics</th>
<th>Economics and Finance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td><strong>First year</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.V. of exam grades</td>
<td>0.113 (0.006)</td>
<td>0.115 (0.004)</td>
<td>0.099 (0.003)</td>
<td>0.109 (0.003)</td>
</tr>
<tr>
<td>Mean exam grade</td>
<td>25.88 (0.459)</td>
<td>25.69 (0.339)</td>
<td>26.70 (0.261)</td>
<td>26.30 (0.232)</td>
</tr>
<tr>
<td><strong>Second year</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.V. of exam grades</td>
<td>0.116 (0.007)</td>
<td>0.119 (0.004)</td>
<td>0.115 (0.002)</td>
<td>0.104 (0.003)</td>
</tr>
<tr>
<td>Mean exam grade</td>
<td>26.35 (0.276)</td>
<td>26.27 (0.207)</td>
<td>26.14 (0.117)</td>
<td>26.77 (0.095)</td>
</tr>
<tr>
<td><strong>Third year</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.V. of exam grades</td>
<td>0.110 (0.008)</td>
<td>0.109 (0.008)</td>
<td>0.123 (0.008)</td>
<td>0.108 (0.002)</td>
</tr>
<tr>
<td>Mean exam grade</td>
<td>26.87 (0.303)</td>
<td>26.93 (0.277)</td>
<td>26.33 (0.298)</td>
<td>26.88 (0.155)</td>
</tr>
</tbody>
</table>

Exams are graded on a scale 0 to 30, pass = 18.

### Table 4: Time together and academic outcomes

<table>
<thead>
<tr>
<th></th>
<th>(log) Mean exam grade$^1$</th>
<th>(log) C.V. of exam grades$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(log) hours in the same class [lnH]</td>
<td>-0.161** (0.070)</td>
<td>-0.154*** (0.044)</td>
</tr>
<tr>
<td>[lnH] x [lnS]</td>
<td>0.036** (0.015)</td>
<td>0.034*** (0.010)</td>
</tr>
<tr>
<td>(log) size of the group [lnS]</td>
<td>-0.255*** (0.095)</td>
<td>-0.240*** (0.060)</td>
</tr>
<tr>
<td>Year dummies x Degree dummies</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Students’ fixed effects</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Observations</td>
<td>3,459</td>
<td>3,459</td>
</tr>
<tr>
<td>Number of students</td>
<td>1,153</td>
<td>1,153</td>
</tr>
</tbody>
</table>

1. Both the coefficient of variation and the mean grade are normalized within degree program and academic year cells. Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Appendix A.   Proofs

Proof of Proposition 1.

Proof. 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]:

\[
\begin{align*}
    f(e_j) - \frac{\gamma}{1 - \epsilon} f(e_i) &> f(e_i) - \gamma(1 - \epsilon) f(e_j) \tag{A.1} \\
    f(e_j) [1 + \gamma (1 - \epsilon)] &> f(e_i) \left[ 1 + \frac{\gamma}{1 - \epsilon} \right] \tag{A.2}
\end{align*}
\]

For equations A.1 and A.2 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

\[
1 + \gamma (1 - \epsilon) > 1 + \frac{\gamma}{1 - \epsilon} \quad 1 - \epsilon > \frac{1}{1 - \epsilon} \tag{A.3}
\]

which is impossible with \( \epsilon < 1 \). ■

Proof of proposition 2.

Proof. Subtracting the first order conditions in equations [10] and [11] from each others yields:

\[
x_i - x_j = \frac{1}{1 + \gamma} \left[ \frac{\partial f(e_i)}{\partial e_i} - (1 - \epsilon) \frac{\partial f(e_j)}{\partial e_j} \right] > 0 \tag{A.4}
\]

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

Proof for proposition 3.

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:

\[
(1 - \epsilon) \left[ \frac{\partial f(e_j)}{\partial e_j} - \frac{\partial^2 f(e_j)}{\partial^2 e_j} \right] \frac{de_j}{d\gamma} = f(e_i) + \gamma \frac{\partial f(e_i)}{\partial e_i} \frac{de_i}{d\gamma} \tag{A.5}
\]

\[
(1 - \epsilon) \left[ \frac{\partial f(e_i)}{\partial e_i} - \frac{\partial^2 f(e_i)}{\partial^2 e_i} \right] \frac{de_i}{d\gamma} = (1 - \epsilon) f(e_j) + \gamma (1 - \epsilon) \frac{\partial f(e_j)}{\partial e_j} \frac{de_j}{d\gamma} \tag{A.6}
\]
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:

\[
\begin{align*}
\left[ \frac{\partial f(e_i)}{\partial e_i} - \frac{\partial^2 f(e_i)}{\partial^2 e_i} \right] \frac{de_i}{d\gamma} &= (1 - \epsilon) + \frac{\gamma (1 - \epsilon) \frac{\partial f(e_j)}{\partial e_j} \frac{f(e_i)}{\partial e_i}}{\frac{\partial f(e_j)}{\partial e_j} - \frac{\partial^2 f(e_j)}{\partial^2 e_j}} + \\
&+ \frac{\gamma^2 (1 - \epsilon) \frac{\partial f(e_j)}{\partial e_j} \frac{\partial f(e_i)}{\partial e_i} \frac{de_i}{d\gamma}}{\frac{\partial f(e_j)}{\partial e_j} - \frac{\partial^2 f(e_j)}{\partial^2 e_j}} \quad (A.7)
\end{align*}
\]

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

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

\[
\frac{\partial f(e_i)}{\partial e_i} \left[ 1 - \gamma^2 (1 - \epsilon) \left( \frac{\partial f(e_j)}{\partial e_j} - \frac{\partial^2 f(e_j)}{\partial^2 e_j} \right) \right] - \frac{\partial^2 f(e_j)}{\partial^2 e_j} > 0 \quad (A.8)
\]

which is in fact positive. ■

**Proof of proposition**

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

\[
\frac{dCV(x)}{d\gamma} = \frac{2}{(x_i + x_j)^2} \left[ x_j \frac{dx_i}{d\gamma} - x_i \frac{dx_j}{d\gamma} \right] = \frac{2(1 - \epsilon)}{(x_i + x_j)^2} \left[ f(e_j) \frac{\partial f(e_i)}{\partial e_i} \frac{de_i}{d\gamma} - f(e_i) \frac{\partial f(e_j)}{\partial e_j} \frac{de_j}{d\gamma} \right] \quad (A.9)
\]

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 \( \frac{de_i}{d\gamma} > \frac{de_j}{d\gamma} \) and the higher the difference between \( \frac{de_i}{d\gamma} \) and \( \frac{de_j}{d\gamma} \) the more likely that \( \frac{dCV(x)}{d\gamma} < 0 \). In particular, it can be shown that if \( \frac{\partial^3 f(e_i)}{\partial^3 e_i} > 0 \) 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 \( \frac{\partial^3 f(e_i)}{\partial^3 e_i} \) must be positive and large enough:

\[
\frac{\partial^3 f(e_i)}{\partial^3 e_i} > 0 \Rightarrow \frac{de_j}{d\gamma} > \frac{de_i}{d\gamma} \Rightarrow \frac{dCV(x)}{d\gamma} > 0 \quad (A.10)
\]
Proof of proposition 5

**Proof.** i) 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 \( e_i^D \):

\[
\frac{\partial^2 g(e_i, e_j)}{\partial^2 e_i} = \frac{\partial g(e_i, e_j)}{\partial e_i} - \gamma (1 - \epsilon) \left( \frac{\partial^2 g(e_j^C, e_i^C)}{\partial e_i^C} \right) (e_i^D - e_i^C) g
\]

(A.11)

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 \( e_j^C < e_j^D \). □

ii) given i) and the complementarity assumption.

Proof of proposition 6

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

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

(A.12)

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 \( e_j^C \):

\[
d(x_i^C - x_j^C) = \frac{\partial^2 g(e_i^C, e_j^C)}{\partial^2 e_i^C} \, de_i + \frac{\partial^2 g(e_i, e_j)}{\partial e_i \partial e_j} \, de_j -
- (1 - \epsilon) \frac{\partial^2 g(e_j^C, e_i^C)}{\partial^2 e_j^C} \, de_j - (1 - \epsilon) \frac{\partial^2 g(e_j^C, e_i^C)}{\partial e_j \partial e_i} \, de_i
\]

\[
= de_i \left[ \frac{\partial^2 g(e_i^C, e_j^C)}{\partial^2 e_i^C} - (1 - \epsilon) \frac{\partial^2 g(e_j^C, e_i^C)}{\partial e_j \partial e_i} \right] -
- de_j \left[ \frac{\partial^2 g(e_i, e_j)}{\partial e_i \partial e_j} - (1 - \epsilon) \frac{\partial^2 g(e_j^C, e_i^C)}{\partial^2 e_j^C} \right] < 0
\]

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

Proof of proposition 7

36
Proof. Simply comparing the first order conditions for the insurance model in equation 19 and the decentralization model in equations 10 or 11 shows that $e^I < e^D$. It, then, follows that $E(x^I) < E(x^D) < E(x^C)$. Under full insurance the performance of both students is identical, hence $CV(x^I) = 0$. ■
## Table B.1: Academic structure

<table>
<thead>
<tr>
<th>Degree program</th>
<th>Course year</th>
<th>Courses</th>
<th>Hours</th>
<th>Subject area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management</td>
<td>First year</td>
<td>Management I</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accounting I</td>
<td>48</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Management II</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Microeconomics</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mathematics</td>
<td>80</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private Law</td>
<td>64</td>
<td>Law</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public Law</td>
<td>32</td>
<td>Law</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Economic History</td>
<td>48</td>
<td>Other</td>
</tr>
<tr>
<td>Second year</td>
<td></td>
<td>Accounting II</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public management</td>
<td>32</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Organization theory</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Macroeconomics</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Statistics</td>
<td>64</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mathematics for finance</td>
<td>32</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commercial Law</td>
<td>64</td>
<td>Law</td>
</tr>
<tr>
<td>Third year</td>
<td></td>
<td>Marketing</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Innovation management</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Corporate finance</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Managerial accounting</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Management of information systems</td>
<td>32</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Strategie management</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Economics of the financial markets</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public Economics</td>
<td>48</td>
<td>Economics</td>
</tr>
<tr>
<td>Economics</td>
<td>First year</td>
<td>Management I</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accounting</td>
<td>48</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Microeconomics</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mathematics</td>
<td>80</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private Law</td>
<td>64</td>
<td>Law</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private Law</td>
<td>32</td>
<td>Law</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Economic history</td>
<td>48</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sociology</td>
<td>48</td>
<td>Other</td>
</tr>
<tr>
<td>Second year</td>
<td></td>
<td>Management II</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Economics of the financial markets</td>
<td>48</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Economic analysis</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Macroeconomics</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mathematics for economics</td>
<td>64</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Statistics</td>
<td>64</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commercial law</td>
<td>64</td>
<td>Law</td>
</tr>
<tr>
<td>Third year</td>
<td></td>
<td>Public economics</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>International economic policy</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data analysis</td>
<td>64</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Econometrics</td>
<td>64</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Management I</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td>Economics and</td>
<td>First year</td>
<td>Accounting</td>
<td>48</td>
<td>Management</td>
</tr>
<tr>
<td>Finance</td>
<td></td>
<td>Economics of financial intermediation</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Microeconomics</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mathematics</td>
<td>80</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private Law</td>
<td>64</td>
<td>Law</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public Economics</td>
<td>32</td>
<td>Law</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Economic History</td>
<td>48</td>
<td>Other</td>
</tr>
<tr>
<td>Second year</td>
<td></td>
<td>Securities market</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Macroeconomics</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monetary economics</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public Economics</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Statistics</td>
<td>64</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mathematics for finance</td>
<td>64</td>
<td>Quantitative subjects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commercial Law</td>
<td>64</td>
<td>Law</td>
</tr>
<tr>
<td>Third year</td>
<td></td>
<td>Management II</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Corporate finance</td>
<td>64</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>International monetary economics</td>
<td>48</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Applied economics</td>
<td>48</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Global banking</td>
<td>48</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Banking</td>
<td>64</td>
<td>Economics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Financial market law</td>
<td>48</td>
<td>Law</td>
</tr>
</tbody>
</table>
Table 5: The crossing property
Dependent variable = 1 if exam grades of students in the pair cross

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours in the same class (/1000) [H]</td>
<td>-0.004**</td>
<td>0.004</td>
<td>-0.003*</td>
<td>0.004*</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>[H] squared</td>
<td>-</td>
<td>-0.008***</td>
<td>-</td>
<td>-0.008***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>Year dummies</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Degree dummies</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Students pair fixed effects</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Observations</td>
<td>313,008</td>
<td>313,008</td>
<td>313,008</td>
<td>313,008</td>
</tr>
<tr>
<td>Number of student pairs</td>
<td>104,336</td>
<td>104,336</td>
<td>104,336</td>
<td>104,336</td>
</tr>
</tbody>
</table>

Each observation is a pair of students in the same degree program in a given academic year.
Standard errors clustered at the student pair level in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%