IDENTIFICATION OF SOCIAL INTERACTIONS THROUGH PARTIALLY OVERLAPPING PEER GROUPS

Giacomo De Giorgi[1] Michele Pellizzari[2] Silvia Redaelli[3]

[1]Stanford University

[2]IGIER-Bocconi University

[3]World Bank

American Economic Journal: Applied Economics,
2 (April 2010): 241-275

MOTIVATION

- Schooling choices are often inefficient:
  - slow adjustment of labour supply to changes in demand (Golding and Katz, 2007)
  - mismatch between workers' skills and and jobs' requirements is widespread (Farber, 1999; Gottschalk and Hansen, 2003; Robst, 2007; Bender and Heywood 2006)
- Adolescents' decisions are heavily influenced by peers' behavior
  - acquisition of information, reference groups, et. (Akerlof and Kranton, 2002)

$\Rightarrow$ **We study the role of peer effects in schooling choices**

WHAT WE DO IN THIS PAPER

- Use data on Bocconi undergraduate students to estimate peer effects in the choice of major
  - we find sizeable and statistically significant effects.
- Estimate the cost in terms of academic and labor market performance of conforming to peers behavior against one's revealed ability
  - we find a significant negative effect on both entry wages and the probability of being mismatched.

THE METHODOLOGICAL CONTRIBUTION

- Identification of peer-effects is problematic:
  1. **Endogeneity**
  2. **Reflection**

- Our strategy uses repeated random allocations into teaching classes:
  1. **Randomisation** solves endogeneity due to self-selection
  2. **Repeated allocation** generates individual-specific groups and solves reflection
  3. **Instrumental variables** solve endogeneity due to unobservable correlated effects.

THE CLASSICAL REFLECTION PROBLEM

- The common specification (Manksy, 1993):

$$y_i = \alpha + \underbrace{\beta E(y|G_i)}_{\text{endogenous effect}} + \underbrace{\gamma E(\mathbf{x}|G_i)}_{\text{exogenous effects}} + \underbrace{\theta \mathbf{z}_g}_{\text{correlated effects}} + \delta \mathbf{x}_i + u_i$$

- Taking the average over group G_i:

$$E(y|G_i) = \left(\frac{\alpha}{1-\beta}\right) + \left(\frac{\gamma+\delta}{1-\beta}\right) E(\mathbf{x}|G_i) + \left(\frac{\theta}{1-\beta}\right) \mathbf{z}_g$$

$\Rightarrow$ the endogenous effect $E(y|G_i)$ is a linear combination of the other regressors

$\Rightarrow$ impossible to separate endogenous, exogenous and correlated effects

THE METHODOLOGICAL CONTRIBUTION

- Identification of peer-effects is problematic:
  1. **Endogeneity**
  2. **Reflection**

- Our strategy uses repeated random allocations into teaching classes:
  1. **Randomisation** solves endogeneity due to self-selection
  2. **Repeated allocation** generates individual-specific groups and solves reflection
  3. **Instrumental variables** solve endogeneity due to unobservable correlated effects.

RELATED PAPERS

- Estimation of peer-effects in education
  - Ammermueller and Pischke (2006), Calvo-Armengol et al. (2006), Carrell et al. (2007, 2009), Cipollone and Rosolia (2007), Hanushek et al. (2003), Hoxby (2000), Sacerdote (2001), Zimmerman (2003).
- Peer-effects in other settings
  - Bayer et al. (2004), Duflo and Saez (2002), Falk and Ichino (2006), Mas and Moretti (2006).
- Methodology
  - Manski (1993), Bramoullè et al. (2006), Brock and Durlauf (2004).

OUTLINE OF THE PRESENTATION

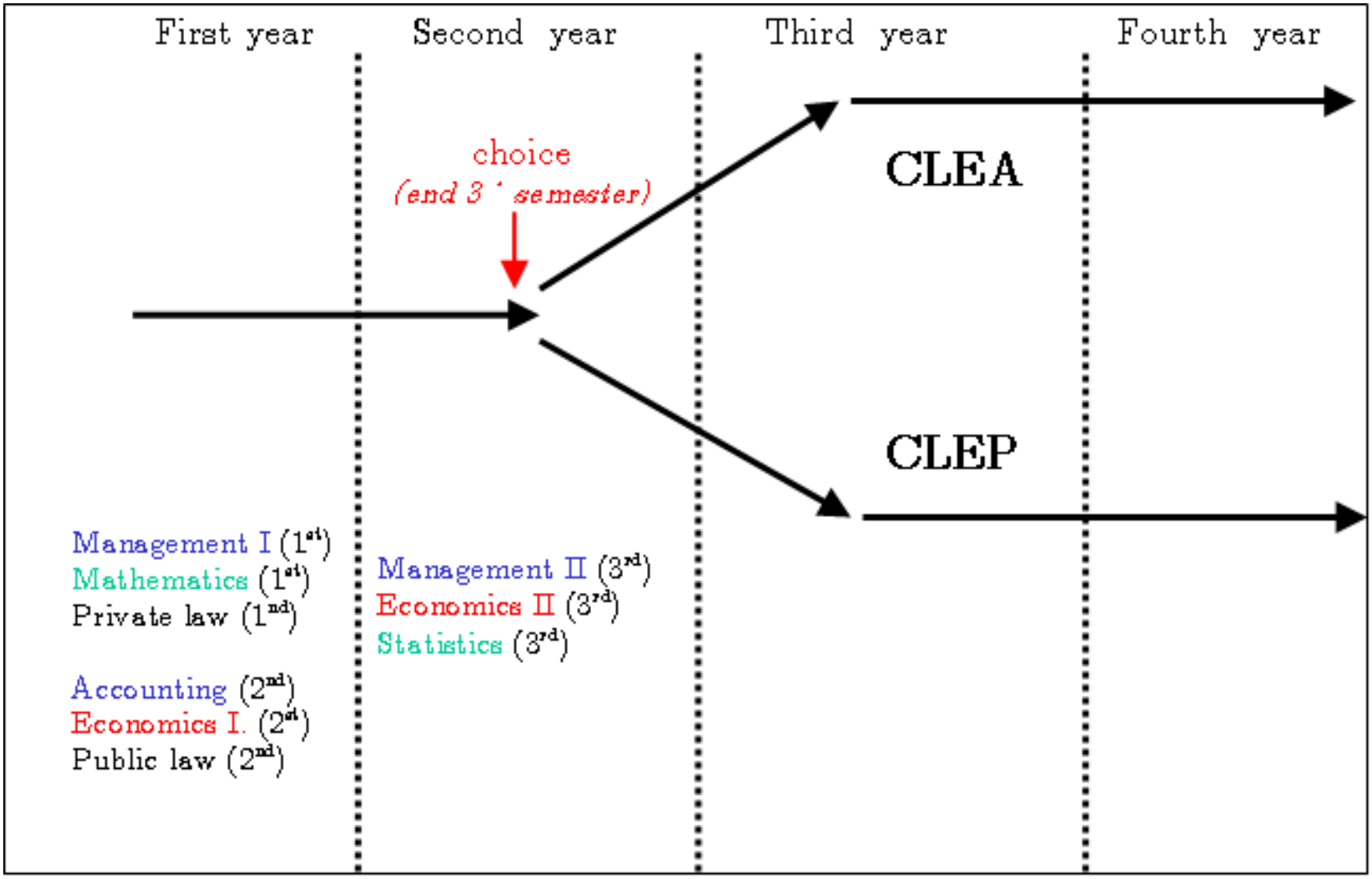
- Introduction
- Bocconi in the CLEA/CLEP era
- The data
- Definition of the peer-groups
- Identification: reflection and endogeneity
- Results
  - Peer-effects in major choice
  - Peer-driven choices and academic and labor market outcomes
- Conclusions

BRIEF OVERVIEW OF THE CLEA/CLEP PROGRAM

- Until the academic year 1999/2000, the most popular Bocconi degree was CLEA/CLEP (started in 1970)
  - official duration was 4 years (8 semesters)
  - 9 common and compulsory courses in first year and a half (3 semesters)
  - then, students choose either Economics (CLEP) or Management (CLEA)
- First available year with reliable information on classes is 1998/1999.
- From academic year 1999/2000, big reform and abolition of CLEA/CELP.

$\Rightarrow$ We focus only on the 1998/1999 cohort of CLEA/CLEP students.

THE CLEA/CLEP TRACK



BRIEF OVERVIEW OF THE CLEA/CLEP PROGRAM

- Until the academic year 1999/2000, the most popular Bocconi degree was CLEA/CLEP (started in 1970)
  - official duration was 4 years (8 semesters)
  - 9 common and compulsory courses in first year and a half (3 semesters)
  - then, students choose either Economics (CLEP) or Management (CLEA)
- First available year with reliable information on classes is 1998/1999.
- From academic year 1999/2000, big reform and abolition of CLEA/CELP.

$\Rightarrow$ We focus only on the 1998/1999 cohort of CLEA/CLEP students.

TEACHING CLASSES

- Students were randomly allocated to teaching classes for each course.
- 9 random allocations during the first three semesters.
- Number of classes varies for each course depending on the number of available teachers.
- The size of the classes varies (within and across courses) because of variation in the physical capacity of the classrooms.
- Randomization designed to fill all classrooms at the same rate.

COURSES AND CLASSES

Course	Sem.	# classes	Average class size	Standard deviation
Management I	1st	10	140.4	14.9
Mathematics	1st	10	140.8	16.9
Private Law	1st	4	351.7	164.1
Accounting	2nd	10	142.8	47.7
Economics I	2nd	6	216.5	92.7
Public Law	2nd	4	351.7	147.8
Economic II	3rd	6	222.8	99.2
Management II	3rd	8	184.2	104.1
Statistics	3rd	8	272.2	90.0

DESCRIPTION OF THE DATA

- Academic records:
  - demographic information: gender, high school type and grade, residence, family income;
  - exam grades and dates, degree program and specialization, graduation date and mark.
- Admission procedures:
  - entry tests, rankings, preferences over the degrees.
- Teaching classes
- Labor market outcomes of graduates in the years 2000-2006.
- Students' evaluations (at the class level).

CHARACTERISTICS OF CLEA/CLEP STUDENTS

	Obs.	High school grade	Admission test
Total	1141	86.3	69.06
CLEP	145	92.2	72.48
CLEA	996	85.4	68.57
Difference (CLEP-CLEA)		6.79***	3.91***

	Av. grades common exams			
	Business	Economics	Quantitative	Total
Total	25.63	24.69	23.67	24.83
CLEP	26.82	26.79	25.81	26.52
CLEA	25.48	24.39	23.35	24.59
Difference (CLEP-CLEA)	1.36***	2.40***	2.46***	1.94***

DEFINITION OF THE PEER-GROUPS

- We define peer groups on the basis of the randomly allocated teaching classes.
- The groups are meant to capture peers with whom students interact socially and academically.
- However, the classes are relatively big and interactions may be limited:
  - $\Rightarrow$ **we eliminate the two law courses** (hence 7 common courses);
  - $\Rightarrow$ **we weight peers by the number of courses taken together**;
  - $\Rightarrow$ **we also produce results restricting peers only to those who have taken at least 4 courses together**.

CHARACTERISTICS OF THE PEER-GROUPS

		Size of peer-groups	
		All peers	Restricted peers
Raw group size	*mean*	647.47	18.08
	std. dev.	79.10	6.77
Average number of classes taken together	*mean*	1.57	4.16
	std. dev.	0.063	0.11
Weighted group size	*mean*	151.07	10.77
	std. dev.	19.73	4.08

PEERS AND LATER ACADEMIC OUTCOMES

Definition of peers:	number of courses attended in the same class...			
	at least 1	at least 2	at least 3	at least 4
	[1]	[2]	[3]	[4]
Panel A: percentage of students who graduate in the same session				
Peers	13.438	13.890	13.346	22.418
Non-peers			12.523	
Diff.	*0.915****	*1.367****	*3.823****	*9.895****
Panel B: percentage of students with the same thesis supervisor				
Peers	1.138	1.150	1.194	1.255
Non-peers			0.957	
Diff.	*0.181****	*0.193****	*0.237****	*0.298****

IDENTIFICATION: REFLECTION AND ENDOGENEITY

1. **Reflection:**
   - In our setting the groups are individual specific
   - $\Rightarrow$ ***no reflection***

2. **Endogeneity:**
   - 2.1 **Self-selection:**
     - in our setting the groups are random
     - additionally, we have very good measures of ability

     $\Rightarrow$ ***no endogeneity***
   - 2.2 **Correlated effects:** instrumental variables

REPEATED RANDOMIZATION SOLVES REFLECTION

Example:

- Students A and B attend 5 courses in the same classes.
- Student B attends some of the remaining 2 courses with student C but not with A.

$\Rightarrow$ $G_A = \{B, +\text{other students}\}$; $G_B = \{A, C, +\text{other students}\}$

$\Rightarrow$ $G_B \neq G_A$ even if all other students are the same

Formally, this implies that $E(y|G_i)$ is **NOT** a linear combination of the other regressors any more:

$$\begin{aligned} E(y_i|G_i) &= \alpha + \beta E[E(y|G_j)|G_i] + \gamma E[E(\mathbf{x}|G_j)|G_i] \\ &\quad + \theta E(\mathbf{z}_{ig}|G_i) + \delta E(\mathbf{x}_i|G_i) \end{aligned}$$

IDENTIFICATION: REFLECTION AND ENDOGENEITY

1. **Reflection:**
   - In our setting the groups are individual specific
   - $\Rightarrow$ ***no reflection***

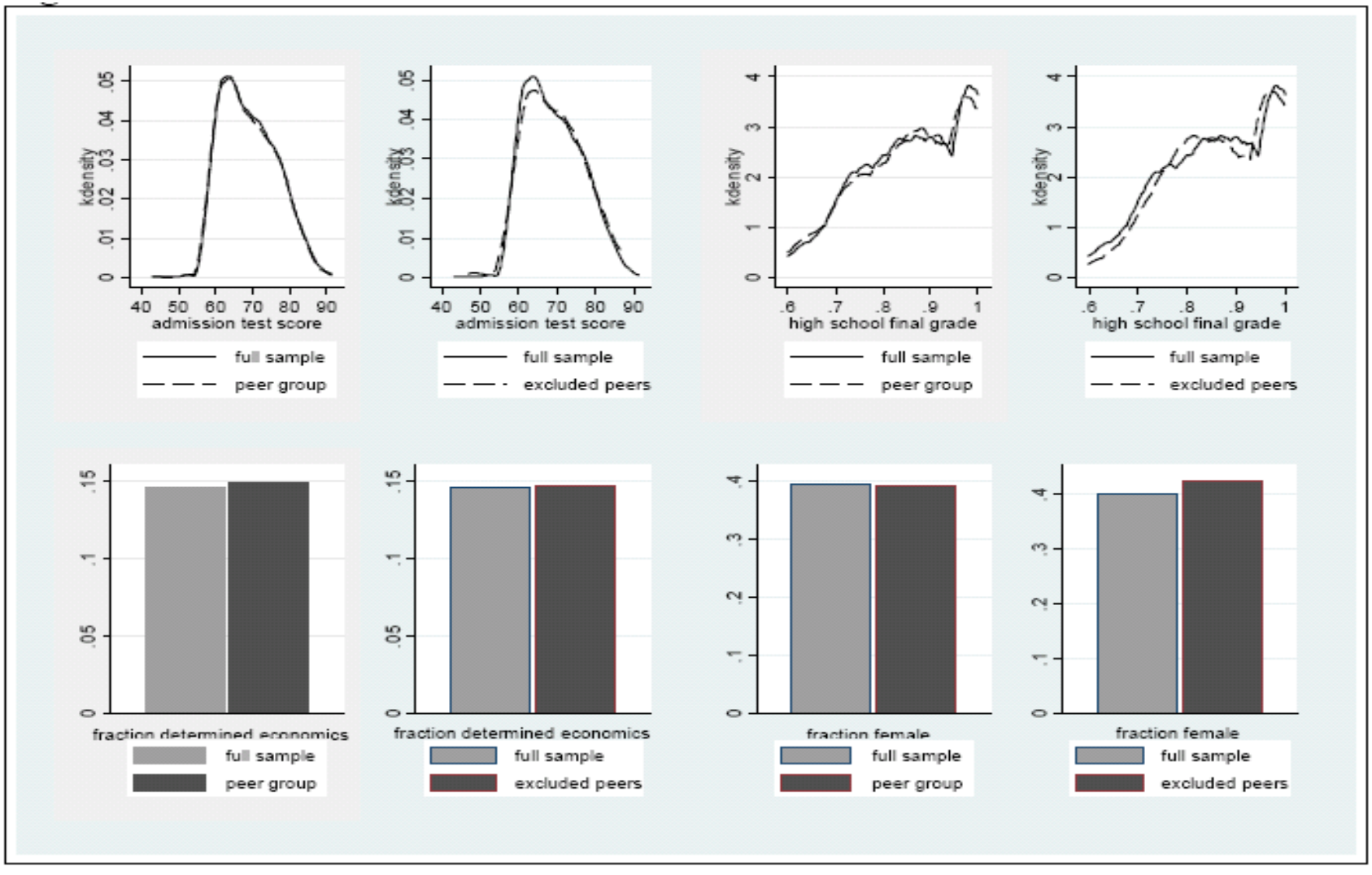
2. **Endogeneity:**
   - 2.1 **Self-selection:**
     - in our setting the groups are random
     - additionally, we have very good measures of ability

     $\Rightarrow$ ***no endogeneity***
   - 2.2 **Correlated effects:** instrumental variables

CORRELATION BETWEEN INDIVIDUAL AND GROUP CHARACTERISTICS

	All peers	Restricted peers
Admission test score	0.0052	0.0236
High school final grade	-0.0325	- 0.0701
Determined to economics	0.0181	0.0169

OBSERVABLE CHARACTERISTICS OF PEERS AND NON-PEERS



IDENTIFICATION: REFLECTION AND ENDOGENEITY

1. **Reflection:**
   - In our setting the groups are individual specific
   - $\Rightarrow$ ***no reflection***

2. **Endogeneity:**
   - 2.1 **Self-selection:**
     - in our setting the groups are random
     - additionally, we have very good measures of ability

     $\Rightarrow$ ***no endogeneity***
   - 2.2 **Correlated effects:** instrumental variables

IV FOR CORRELATED EFFECTS

- Group unobservable shocks (teachers, classrooms, et.) may induce endogeneity of $E(y|G_i)$
- The exogenous characteristics of EXCLUDED PEERS are natural valid instruments:
  - from the previous example: $B \in G_A$; $C \in G_B - G_A$.
  - $\Rightarrow$ use the $\mathbf{x}_C$'s as instruments for y_B in the equation for y_A.
- By construction, the $\mathbf{x}_C$'s are:
  - UNCORRELATED with the group fixed effect of A;
  - CORRELATED with y_B through endogenous interactions: $\mathbf{x}_C \Rightarrow y_C \Rightarrow$ (via endogenous interactions) y_B

IDENTIFICATION: REFLECTION AND ENDOGENEITY

1. **Reflection:**
   - In our setting the groups are individual specific
   - $\Rightarrow$ ***no reflection***

2. **Endogeneity:**
   - 2.1 **Self-selection:**
     - in our setting the groups are random
     - additionally, we have very good measures of ability

     $\Rightarrow$ ***no endogeneity***
   - 2.2 **Correlated effects:** instrumental variables

THE SOURCES OF EMPIRICAL VARIATION

- Where does the variation in $\overline{y}_i$ comes from?
  1. Different combinations of the class shocks, i.e. the z_i's are different for most peers;
  2. sample variation in $\overline{\mathbf{x}}_i$, especially with small classes:
     - randomization implies little variation in $\overline{\mathbf{x}}_i$ in large groups, so exogenous effects can be identified separately from the constant only with restricted groups.
  3. Possibly non-linear combinations of individual (and group) shocks within groups (not explicitly modeled).

OUTLINE OF RESULTS

Two sets of empirical results:

1. Estimate endogenous peer-effects in the choice of major
2. Define four decision modes and estimate their effect on:
   - academic outcomes
   - labor market outcomes (wages and mismatch)

1. PEER-EFFECTS IN THE CHOICE OF MAJOR

Eventually, we estimate the following model:

$$y_i = \alpha + \beta \overline{y}_i + \gamma \overline{\mathbf{x}}_i + \delta \mathbf{x}_i + u_i$$

where:

- $y_i = 1$ if student i chooses economics, 0 otherwise
- $\overline{y}_i$=(weighted) % of peers choosing economics
- $\overline{\mathbf{x}}_i$=(weighted) peers' characteristics
- $\mathbf{x}_i$=controls
  - ability measures (high-school grades, entry test score)
  - individual characteristics (gender, household income, residence, preference for economics)
- 2SLS estimation:
  - Excluded instruments: excluded peers' averages of admission test, high school final grade, preference for economics

1. PEER-EFFECTS IN THE CHOICE OF MAJOR

Effect of 1 additional *average* peer choosing CLEP ($\Delta \overline{y}_i = 1$)

	OLS	2SLS
Restricted peers:		
with $\overline{\mathbf{x}}_i$	0.008	0.074*
	[0.184]	[0.054]
First-stage F-test		10.84
without $\overline{\mathbf{x}}_i$	0.001	0.069*
	[0.155]	[0.053]
First-stage F-test		12.17
All peers:		
without $\overline{\mathbf{x}}_i$	0.001	0.008*
	[0.193]	[0.068]
First-stage F-test		30.30
Obs.	1141	1141

Asymptotic p-values in brackets

INTERPRETATION OF THE EFFECTS

- Our preferred specification is *IV-restricted peers*.
- In this specification 1 additional average peer who chooses CLEP increases one's probability of choosing the same major by 7.4 percentage points.
- Over an average of about 13%, this is an increase of about 57%.
- The average weighted group size is about 10 and among these there is on average 1.3 who choose CLEP, so this is the effect of almost doubling the share of one's peers who choose CLEP
- In other words, a 10% increase in the share of peers who choose CLEP increases one's probability of taking the same major by approximately 6%.

ROBUSTNESS CHECKS

1. Groups based on 5 courses (only economics and business)
2. Exponential weighting scheme
3. Placebo peer groups
4. Teacher quality controls
5. Course congestion

ROBUSTNESS CHECKS

	Restricted peers	
	OLS	2SLS
Groups based on 5 courses	0.011 [.137]	0.108* [0.060]
Exponential weights	0.006 [0.279]	0.074* [0.069]
Placebo groups	-0.006 [0.210]	-0.022 [0.720]
Teacher quality dummies	0.004 [0.494]	0.068* [0.096]
Course congestion	0.008 [0.171]	0.073* [0.058]

Asymptotic p-values in brackets

OUTLINE OF RESULTS

Two sets of empirical results:

1. Estimate endogenous peer-effects in the choice of major
2. Define four decision modes and estimate their effect on:
   - academic outcomes
   - labor market outcomes (wages and mismatch)

2. DEFINITION OF DECISION MODES

- We define 4 decision modes on the basis of 2 indicators:
- First indicator:

$$f_i = \begin{cases} \dfrac{\sum_{j \in G_i} \omega_j ECON_j}{N^{-1} \sum ECON_j} & \text{if } ECON_i = 1 \\ \dfrac{\sum_{j \in G_i} \omega_j BUSIN_j}{N^{-1} \sum BUSIN_j} & \text{if } BUSIN_i = 1 \end{cases}$$

- Second indicator:

$$g_i = \begin{cases} \dfrac{GPA_i^{ECON}}{GPA_i^{BUSINESS}} \cdot \dfrac{\sum GPA_j^{BUSINESS}}{\sum GPA_j^{ECON}} & \text{if } ECON_i = 1 \\ \dfrac{GPA_i^{BUSINESS}}{GPA_i^{ECON}} \cdot \dfrac{\sum GPA_j^{ECON}}{\sum GPA_j^{BUSINESS}} & \text{if } BUSIN_i = 1 \end{cases}$$

2. DISTRIBUTION OF DECISION MODES

	$g_i > 1$	$g_i < 1$
$f_i > 1$	Coherent 29.11%	Peer driven 27.56%
$f_i < 1$	Ability driven 23.11%	Incoherent 20.16%

(based on restricted peers)

2. ARE BOOKS BETTER THAN COMPANY?

- Estimate the effect of decision modes on academic outcomes:

  $$y_i = c + \pi_1[\text{peer driven}]_i + \pi_2[\text{coherent}]_i + \pi_3[\text{incoherent}]_i + \vartheta \mathbf{x}_i + u_i$$

  with the *ability driven* as a reference group

- y_i= average grade (non. common), graduation mark, time to graduation
- Identification requires that, conditional on the observable $\mathbf{x_i}$ the decision modes are exogenous
  - $\mathbf{x_i}$ includes detailed controls for ability (high-school grades and type, average grades, ratio between average grade in economics and business courses, number of exams taken on the first available session)

2. DECISION MODES AND ACADEMIC OUTCOMES

	Average grade (non-com. exams)		Graduation mark		Time to graduation
Peer driven	**-0.205****	**-0.168***	**-0.737****	**-0.641****	**0.060**
	(0.091)	**(0.088)**	**(0.308)**	**(0.304)**	**(0.053)**
Coherent	-0.058	-0.050	-0.330	-0.310	0.006
	(0.090)	(0.088)	(0.301)	(0.297)	(0.052)
Incoherent	-0.217**	-0.123	-0.776**	-0.534	0.182***
	(0.102)	(0.099)	(0.334)	(0.333)	(0.059)
Av. grade	0.643***	0.595***	2.875***	2.749***	-0.168***
	(0.020)	(0.021)	(0.070)	(0.072)	(0.014)
Time to graduation	–	-0.479***	–	-1.245***	–
		(0.054)		(0.180)	
Admission test	0.006	0.005	-0.021	-0.023	0.001
	(0.005)	(0.005)	(0.016)	(0.016)	(0.003)

2. LABOUR MARKET OUTCOMES

- Surveys of Bocconi graduates at 1.5 years since graduation:
  - surveys cover all graduates between 2000 and 2006
  - low response rates (about 43% matched on average in our cohort) but mostly random or driven by observables (survey wave, gender, residence)
- Labour market outcomes:
  - wage in the first job in euros at 2005 prices (recorded in intervals)
  - mismatch = having encountered difficulties in the first job (tasks too difficult, relational problems, jobs does not fit personal attitudes, et.)

2. DECISION MODES AND LABOUR MARKET OUTCOMES

	Entry wage		1=mismatched	
Peer driven	**-0.130**	**-0.136***	**0.443****	**0.438****
	(0.081)	**(0.081)**	**(0.197)**	**(0.199)**
m.eff.	-0.130	-0.136	*0.135*	*0.133*
Coherent	-0.153**	-0.141**	0.241	0.242
	(0.074)	(0.076)	(0.180)	(0.180)
Incoherent	-0.082	-0.096	0.362*	0.376**
	(0.084)	(0.086)	(0.190)	(0.191)
Graduation mark	-	0.018***	-	-0.002
		(0.006)		(0.013)
Time to graduation	-	0.054	-	-0.410***
		(0.097)		(0.150)
Observations	427	427	436	436

CONCLUSIONS

- New estimation strategy for the identification of endogenous and exogenous peer effects separately
- The new strategy is very general:
  - in the real world reference groups overlap and rarely coincide perfectly;
  - perhaps stricter data requirements...
  - ...but some existing datasets could already be exploited with our approach
    (Bayer et al., 2004; Calvó et al. ,2006; Mas and Moretti, 2006)

CONCLUSIONS (2)

- Identifying endogenous social interactions is important for policy!
  1. the magnitude of the effect of manipulations of exogenous/group characteristics depends on the extent of endogenous interactions;
  2. large-scale interventions with general equilibrium effects require knowledge of the endogenous interactions parameter for proper design and evaluation.